

STIC-Biotech/ChemLib

69910

Fr m:
Sent:
To:
Subject:

Nguyen, Quang (AU1632)
Monday, June 03, 2002 3:14 PM
STIC-Biotech/ChemLib
Sequence search request for 09/721543

Good afternoon,

Please search:

SEQ ID NO:12, 13, 14, 15 and 16 against commercial, issued and pending U.S. application databases.
Should there be questions, please contact me at 308-8339 (CM1-12A12, AU 1636, my mail box is in CM1-11E12).

THANK YOU.

RECEIVED
JUN - 3 2002
(STIC)

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Searcher: Bob
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 6-5-02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:09:09 ; Search time 4272.65 Seconds
(without alignments)
116.880 Million cell updates/sec

Title: US-09-721-543A-16
Perfect score: 37
Sequence: 1 ctatatccaccatattcccttcccttgcgtgtggg 37

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1 | 23 | 62.2 | 513 | 12 | AZ908796 |
| C 2 | 22.8 | 61.6 | 278 | 12 | AZ742846 |
| C 3 | 22.6 | 61.6 | 626 | 12 | AQ348364 |
| C 4 | 22.6 | 61.1 | 564 | 10 | BI682748 |
| C 5 | 22.6 | 61.1 | 934 | 12 | AQ746618 |
| C 6 | 22 | 59.5 | 359 | 9 | AI381292 |
| C 7 | 21.8 | 58.9 | 197 | 9 | AW531399 |
| C 8 | 21.8 | 58.9 | 281 | 10 | BF416687 |
| C 9 | 21.8 | 58.9 | 306 | 9 | AW433771 |
| C 10 | 21.8 | 58.9 | 369 | 9 | AV656658 |
| C 11 | 21.8 | 58.9 | 370 | 9 | AA997381 |
| C 12 | 21.8 | 58.9 | 478 | 10 | BG381507 |
| C 13 | 21.8 | 58.9 | 504 | 9 | AI575981 |
| C 14 | 21.8 | 58.9 | 696 | 9 | AV752617 |
| C 15 | 21.8 | 58.9 | 916 | 10 | BE965057 |
| C 16 | 21.8 | 58.9 | 947 | 10 | BE911323 |
| C 17 | 21.8 | 58.9 | 1275 | 10 | BF688465 |

RESULT 1
AZ908796/c
LOCUS
DEFINITION
ACCESSION
AZ908796
VERSION
AZ908796.1 GI:13227741
KEYWORDS
GSS
SOURCE
house mouse.
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 513)
AUTHORS
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsagay,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE
JOURNAL
COMMENT
Mouse BAC End Sequences from library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html
Plate: 222 row: E column: 16
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .513
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-222E16"
/clone_lib="RPCI-24"

FEATURES
source

C 18 21.6 58.4 288 9 AW242494
C 19 21.6 58.4 376 10 BF226236
C 20 21.6 58.4 523 9 AII151930
C 21 21.6 58.4 544 10 BF286029
C 22 21.6 58.4 600 10 BI990118
C 23 21.6 58.4 683 9 BB140912
C 24 21.6 58.4 698 10 BG923513
C 25 21.6 58.4 762 10 BG620158
C 26 21.6 58.4 764 10 BI688856
C 27 21.6 58.4 815 10 BI645932
C 28 21.6 58.4 822 10 BI522467
C 29 21.4 57.8 214 12 BH626108
C 30 21.4 57.8 265 9 AW353241
C 31 21.4 57.8 348 9 AW336288
C 32 21.4 57.8 379 12 BH628904
C 33 21.4 57.8 442 9 AW478514
C 34 21.4 57.8 445 10 BE753768
C 35 21.4 57.8 474 12 BH625558
C 36 21.4 57.8 475 12 BH408135
C 37 21.4 57.8 497 10 BF039832
C 38 21.4 57.8 519 10 BM365942
C 39 21.4 57.8 543 9 AW657071
C 40 21.4 57.8 574 12 BH631145
C 41 21.4 57.8 577 10 BI540444
C 42 21.4 57.8 1350 12 AG054991
C 43 21.2 57.3 413 12 AZ274408
C 44 21.2 57.3 420 9 AI328516
C 45 21.2 57.3 480 10 BI490098

ALIGNMENTS

AZ908796 513 bp DNA linear GSS 05-MAR-2001
RPCI-24-222E16.TJ RPCI-24 Mus musculus genomic clone RPCI-24-222E16
, DNA sequence.

ACCESSION
AZ908796
VERSION
AZ908796.1 GI:13227741
KEYWORDS
GSS
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse BAC End Sequences from library RPCI-24

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html

Plate: 222 row: E column: 16

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .513

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-222E16"

/clone_lib="RPCI-24"

/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTABAC1; Site_1: BamH1; Site_2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTABAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 158 a 107 c 112 g 136 t

ORIGIN

Query Match 62.2%; Score 23; DB 12; Length 513;

Best Local Similarity 83.9%; Pred. No. 3.9e+02;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 atccaccatcccccttcccccttgcgtgtg 35

||||| ||||| ||||| ||||| ||||| |||||

Db 270 AGCCTCCCATATCCCTTCCTTGCACGTGTG 240

RESULT 2

AZ742846/c

LOCUS

DEFINITION RPCI-24-110L21-TV RPCI-24 Mus musculus genomic clone RPCI-24-110L21

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 278)

Tseng, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,

Russell, D., de Jong, P., and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 110 row: L column: 21

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers

1..278

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-110L21"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTABAC1; Site_1: BamH1; Site_2: BamH1;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTABAC1 cloning vector at the

BamH1 sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 100 a 32 c 80 g 66 t

ORIGIN

Query Match 61.6%; Score 22.8; DB 12; Length 278;

Best Local Similarity 92.3%; Pred. No. 4e+02;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 cccatcccccttcccccttgcgtgtg 35

||||| ||||| ||||| ||||| ||||| |||||

Db 101 CCCATATCCCTTCCCTTCCCTGTGTG 76

RESULT 3

AQ348364/c

LOCUS

DEFINITION RPCI11-137E20-TV RPCI-11 Homo sapiens genomic clone RPCI-11-137E20,

DNA sequence.

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 626)

Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P., and Venter

J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Map Building

Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pdejong@med.bu.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.bu.edu/ordering) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers

1..626

/organism="Homo sapiens"

/db_xref="GDB:752339"

/db_xref="taxon:9606"

/clone="RPCI-11-137E20"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 188 a 162 c 130 g 144 t

ORIGIN

Query Match 61.6%; Score 22.8; DB 12; Length 636;

Best Local Similarity 79.4%; Pred. No. 4.8e+02;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 tatccaccatcccccttcccccttgcgtgtg 37

||||| ||||| ||||| ||||| ||||| |||||

Db 282 TGTCCTTCCCATATTCCTTCTCTTGTGCTGGG 249

RESULT 4

BI682748/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos taurus

BI682748

463908 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.

EST.

GI:15635682

EST.

COW.

Bos taurus

BI682748

564 bp mRNA linear EST 17-SEP-2001

went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 102 a 80 c 108 g 67 t 58.9% Score 22; DB 9; Length 359; Best Local Similarity 78.1%; Pred. No. 7.9e+02; Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Query Match 59.5%; Score 22; DB 9; Length 359; Best Local Similarity 78.1%; Pred. No. 7.9e+02; Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 tccaccatcccccttcccttgcgtggtg 37
Db 314 TCCACCATGCCCATCCATTCATGTGNG 345

RESULT 7
AW531399/c
LOCUS
DEFINITION
UI-R-B50-amm-e-05-0-UI.s1 197 bp mRNA linear EST 06-MAR-2000
UI-R-B50-amm-e-05-0-UI 3', mRNA sequence.

ACCESSION
AW531399
VERSION
AW531399.1 GI:7173813
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus

REFERENCE
AUTHORS
Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
MEDLINE
Genome Res. 6 (9), 791-806 (1996)
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-Yes.

FEATURES
source
1. .197
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B50-amm-e-05-0-UI"
/clone_lib="UI-R-B50"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-B50 library is derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu."
TAG_LIB="UI-R-B50"
TAG_TISSUE="embryo at 13 dpc"
TAG_SEQ="AATCC"
47 a 35 c 62 g 53 t

BASE COUNT
ORIGIN

Query Match 58.9%; Score 21.8; DB 9; Length 197; Best Local Similarity 78.8%; Pred. No. 8.1e+02; Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatatccaccatcccccttcccttgcgtg 33
Db 112 CTATCTCCCCCATGTCCCCATGCCCTTCCCTG 80

RESULT 8
BF416687/c
LOCUS
DEFINITION
UI-R-CNO-bkz-b-11-0-UI.s1 281 bp mRNA linear EST 28-NOV-2000
UI-R-CNO-bkz-b-11-0-UI 3', mRNA sequence.

ACCESSION
BF416687
VERSION
BF416687.1 GI:11404716
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus

REFERENCE
AUTHORS
Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
MEDLINE
Genome Res. 6 (9), 791-806 (1996)
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized fundus library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-21, >AT-rich#Low_complexity
Seq primer: M13 Forward
POLYA-Yes.

FEATURES
source
1. .281
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CNO-bkz-b-11-0-UI"
/clone_lib="UI-R-CNO"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CNO library is a normalized library constructed from the following tissues: brown adipose, penis, salivary glands, bladder, fundus, cervix, seminal vesicles. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB="UI-R-CNO"
TAG_TISSUE="fundus"
TAG_SEQ="TTCCG"
67 a 52 c 86 g 76 t

BASE COUNT
ORIGIN

Query Match 58.9%; Score 21.8; DB 10; Length 281;

| | | |
|-----------------------|------------------|------------------|
| Best Local Similarity | 78.8%; | Pred. No. 9e+02; |
| Matches | 26; Conservative | 0; Mismatches |
| | | 7; Indels |
| | | 0; Gaps |
| | | 0; |

QY 1 ctatatccaccatatacccccttgcgtg 33
||||| ||| ||||| ||||| ||||| |||
Dd 112 CTATCTCCCCCATGTCCCCATGCCCTCCCTG 80

| | |
|------------|--|
| RESULT | 10 |
| AV656658/c | |
| LOCUS | AV656658 |
| DEFINITION | GLC Homo sapiens cDNA clone GLCTF05 3', mRNA sequence. |
| ACCESSION | AV656658 |
| VERSION | AV656658.1 |
| KEYWORDS | EST. |
| SOURCE | human. |

| SOURCE | ORGANISM | Human. | Homo sapiens | Chordata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. |
|-----------|---|--------|--------------|---|
| REFERENCE | Xu Y, Huang J, Yu Z, Qian B, Zhu Z, Yan Q, Cao M, Zhang Y | | | |

AUTHORS
Xiao, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xia, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z., and Han, Z.

TITLE
Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

JOURNAL
MEDLINE
COMMENT

201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

```

/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XbaI"

```

```

/tissue_types="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      101 a      82 c      116 g      70 t
ORIGIN

Query Match      58.9%; Score 21.8; DB 9; Length 369;
Best Local Similarity 78.8%; Pred. No. 9.4e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      3      atatccaccacatatcccccctccccccttgcgtagt 35
      || ||||| ||||| ||||| ||||| ||
Db      335      ATTCCCCACCGGATATCCCGCTTCCCGTTCGAGAAAT 303

```

| | |
|------------|--|
| DEFINITION | UI-R-CO-hq-f-11-0-013' UI-R-CO hatus norvegicus cDNA clone |
| ACCESSION | AA997381 |
| VERSION | AA997381.1 |
| KEYWORDS | EST. |
| SOURCE | Norway rat |

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;


```
/clone="IMAGE:3886043"  
/clone_lib="Nih_MGC_89"  
/tissue_type="large cell carcinoma, undifferentiated"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;  
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."
```

```
BASE COUNT      321 a      167 c      279 g      149 t  
ORIGIN
```

```
Query Match      58.9%; Score 21.8; DB 10; Length 916;  
Best Local Similarity 78.8%; Pred. NO. 1.1e+03;  
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy 5 atccaccatattcccttcccttgcgtgtgg 37  
||||| ||| | ||||| ||||| ||||| |||||  
Db 670 ATCCGCCCTTCTCCCGCCCGCCCTTTGGGTGGG 638
```

```
Search completed: June 4, 2002, 17:09:16  
Job time: 4549 sec
```



```
Qy 5 atccaccatcccccttcgctggtgg 37
Db 339 ATCCACTGTGATGCACCTTCCCTTGGGGTGG 307

RESULT 2
US-08-481-814A-2/C
; Sequence 2, Application US/08481814A
; Patent No. 5869040
; GENERAL INFORMATION:
; APPLICANT: Hsu, Yen-Ming
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Biogen, Inc.
; STREET: 14 Cambridge Center
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02142
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.814A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan, Warren A
; REFERENCE/DOCKET NUMBER: A001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-679-2000
; TELEFAX: 617-679-2838
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1766 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: HeLa
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 429..1739
; OTHER INFORMATION: /product= "E2F-2"
US-08-481-814A-2

Query Match 54.68; Score 20.2; DB 2; Length 1766;
Best Local Similarity 75.88; Pred. No. 17;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 atccaccatcccccttcgctggtgg 37
Db 767 ATCCACTGTGATGCACCTTCCCTTGGGGTGG 735

RESULT 3
US-08-136-119-3/C
; Sequence 3, Application US/08136119
; Patent No. 5473056
; GENERAL INFORMATION:
; APPLICANT: Heimbrosk, David C.
; APPLICANT: Hoyle, Mona I.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: E 2F-2, A NOVEL MAMMALIAN TRANSCRIPTION
; TITLE OF INVENTION: FACTOR
```

```
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Muthard
; STREET: P.O. Box 2000, 126 Lincoln Avenue
; CITY: Rahway
; STATE: N.J.
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136.119
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 19042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2647 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-136-119-3

Query Match 54.68; Score 20.2; DB 1; Length 2647;
Best Local Similarity 75.88; Pred. No. 18;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 atccaccatcccccttcgctggtgg 37
Db 767 ATCCACTGTGATGCACCTTCCCTTGGGGTGG 735

RESULT 4
US-08-845-998-3
; Sequence 3, Application US/08845998
; Patent No. 5879892
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coullie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845.998
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
```

; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..577
US-08-845-998-3

Query Match 52.4%; Score 19.4; DB 2; Length 2259;
Best Local Similarity 79.3%; Pred. No. 36;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattcccttcccttgcgtg 35
||||| | ||||| ||||| |||||
Db 2139 CCACCTGAGCCCTTCCCTTCTTGAG 2167

RESULT 5
US-09-206-537-3
; Sequence 3, Application US/09206537
; Patent No. 6130052
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/206,537
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..577
US-09-206-537-3

Query Match 52.4%; Score 19.4; DB 3; Length 2259;
Best Local Similarity 79.3%; Pred. No. 36;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattcccttcccttgcgtg 35
||||| | ||||| ||||| |||||
Db 2139 CCACCTGAGCCCTTCCCTTCTTGAG 2167

RESULT 6
US-09-430-854-3
; Sequence 3, Application US/09430854
; Patent No. 6271019
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/430,854
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..577
US-09-430-854-3

Query Match 52.4%; Score 19.4; DB 4; Length 2259;
Best Local Similarity 79.3%; Pred. No. 36;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattccctcccttgcgtg 35
||||| | ||||| ||||| |||
Db 2139 CCACCTGAGCCCTCCCTTTCTTGAG 2167

RESULT 7

US-08-756-506-23/c
; Sequence 23, Application US/08756506
; Patent No. 5905185

; GENERAL INFORMATION:

; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; TITLE OF INVENTION: ANIMALS
; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA

; COUNTRY: USA

; ZIP: 98102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/756,506

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Sawislak, Deborah A

; REGISTRATION NUMBER: 37,438

; REFERENCE/DOCKET NUMBER: 95-28

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-442-6672

; TELEFAX: 206-442-6678

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6763 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-756-506-23

Query Match 52.4%; Score 19.4; DB 2; Length 6763;

Best Local Similarity 79.3%; Pred. No. 43;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattccctcccttgcgtg 35
||||| | ||||| ||||| |||

Db 4342 CCACCAAGACCCCTACCCCATGTGTG 4314

RESULT 8

US-08-206-176-7/c

; Sequence 7, Application US/08206176

; Patent No. 5639940

; GENERAL INFORMATION:

; APPLICANT: Garner, Ian

; APPLICANT: Dairymple, Michael A

; APPLICANT: Prunkard, Donna E

; APPLICANT: Foster, Donald C

; TITLE OF INVENTION: Production of Fibrinogen in Transgenic

; TITLE OF INVENTION: Animals

; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,176
; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Parker, Gary E

; REGISTRATION NUMBER: 31-648

; REFERENCE/DOCKET NUMBER: 93-15

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-547-8080 ext 322

; TELEFAX: 206-548-2329

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10807 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: ovine beta-lactoglobulin

; US-08-206-176-7

Query Match 52.4%; Score 19.4; DB 1; Length 10807;

Best Local Similarity 79.3%; Pred. No. 46;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattccctcccttgcgtg 35
||||| | ||||| ||||| |||

Db 8386 CCACCAAGACCCCTACCCCATGTGTG 8358

RESULT 9

US-08-756-506-5/c

; Sequence 5, Application US/08756506

; Patent No. 5905185

; GENERAL INFORMATION:

; APPLICANT: Garner, Ian

; APPLICANT: Cottingham, Ian R.

; APPLICANT: Temperley, Simon M.

; APPLICANT: Foster, Donald C.

; APPLICANT: Sprecher, Cindy A.

; APPLICANT: Prunkard, Donna E.

; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC

; TITLE OF INVENTION: ANIMALS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ZymoGenetics, Inc.

; STREET: 1201 Eastlake Avenue East

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/756,506

; FILING DATE:

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 95-28
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10807 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-756-506-5

Query Match 52.4%; Score 19.4; DB 2; Length 10807;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 ccaccatccccctcccttcgctgtg 35
||||| ||||| ||||| |||||
Db 8386 CCACCAAGACCCCTACCCCATGTGTG 8358

RESULT 10
US-09-613-182-13/c
; Sequence 13, Application US/09613182
; Patent No. 6294653
; GENERAL INFORMATION:
; APPLICANT: Mayfield, Stephen
; TITLE OF INVENTION: RNA BINDING PROTEIN AND BINDING SITE USEFUL FOR
; FILE REFERENCE: SCR2177S
; CURRENT APPLICATION NUMBER: US/09/613,182
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 09/341,550
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: PCT/US98/00840
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 60/035,955
; PRIOR FILING DATE: 1997-01-17
; PRIOR APPLICATION NUMBER: 60/069,400
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1424
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (252)..(1310)
; NAME/KEY: misc_feature
; LOCATION: (279)
; OTHER INFORMATION: Codon also can encode Ser
; NAME/KEY: misc_feature
; LOCATION: (282)
; OTHER INFORMATION: Codon also can encode Glu
; NAME/KEY: misc_feature
; LOCATION: (294)
; OTHER INFORMATION: Codon also can encode Gly
; NAME/KEY: misc_feature
; LOCATION: (306)
; OTHER INFORMATION: Codon also can encode Asn
; NAME/KEY: misc_feature
; LOCATION: (357)
; OTHER INFORMATION: Codon also can encode Leu
; NAME/KEY: misc_feature
; LOCATION: (369)
; OTHER INFORMATION: Codon also can encode Thr
; NAME/KEY: misc_feature

; LOCATION: (486)
; OTHER INFORMATION: Codon also can encode Ser
; NAME/KEY: misc_feature
; LOCATION: (495)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (510)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (555)
; OTHER INFORMATION: Codon also can encode Val
; NAME/KEY: misc_feature
; LOCATION: (588)
; OTHER INFORMATION: Codon also can encode Glu
; NAME/KEY: misc_feature
; LOCATION: (600)
; OTHER INFORMATION: Codon also can encode Leu
; NAME/KEY: misc_feature
; LOCATION: (621)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (714)
; OTHER INFORMATION: Codon also can encode Thr
; NAME/KEY: misc_feature
; LOCATION: (729)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (1092)
; OTHER INFORMATION: Codon also can encode Val
; NAME/KEY: misc_feature
; LOCATION: (1191)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (1284)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (1287)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (1290)
; OTHER INFORMATION: Codon also can encode Glu
; NAME/KEY: misc_feature
; LOCATION: (1293)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (1296)
; OTHER INFORMATION: Codon also can encode Pro
; NAME/KEY: misc_feature
; LOCATION: (1302)
; OTHER INFORMATION: Codon also can encode Thr
; NAME/KEY: misc_feature
; LOCATION: (1308)
; OTHER INFORMATION: Codon also can encode Gly
; US-09-613-182-13

Query Match 51.9%; Score 19.2; DB 4; Length 1424;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 tatatccaccatcccccttcgctgtg 33
||||| ||||| ||||| |||||
Db 117 TTTATGTACCTACGTCCCTTCCCTTACGG 86

RESULT 11
US-08-466-603-4
; Sequence 4, Application US/08466603
; Patent No. 5726018
; GENERAL INFORMATION:
; APPLICANT: Pasternack, Gary R.
; APPLICANT: Kuhnajda, Francis P.
; TITLE OF INVENTION: No. 5726018el Mammalian Protein Associated With

;; TITLE OF INVENTION: Uncontrolled Cell Division
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner, Birch, McKie & Beckett
;; STREET: 1001 G Street, N.W.
;; CITY: Washington, D.C.
;; STATE: District of Columbia
;; COUNTRY: U.S.A.
;; ZIP: 20001
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/466,603
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/314,503
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Posorske Esq., Laurence H.
;; REGISTRATION NUMBER: 34,698
;; REFERENCE/DOCKET NUMBER: 1107.47218
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202 508-9153
;; TELEFAX: 202 508-9299
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 759 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORGANISM: Mus sp
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..548
US-08-466-603-4

Query Match 51.4%; Score 19; DB 1; Length 759;
Best Local Similarity 81.5%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatatccaccatccctccctc 27
||| | ||||| ||||| |||||
Db 591 CTGTTTACCCTATCCCTCCCT 617

RESULT 12
US-08-314-503A-4
;; Sequence 4, Application US/08314503A
;; Patent No. 5734022
;; GENERAL INFORMATION:
;; APPLICANT: Pasternack, Gary R.
;; APPLICANT: Kuhajda, Francis P.
;; TITLE OF INVENTION: No. 5734022el Mammalian Protein Associated With
;; TITLE OF INVENTION: Uncontrolled Cell Division
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner, Birch, McKie & Beckett
;; STREET: 1001 G Street, N.W.
;; CITY: Washington, D.C.
;; STATE: District of Columbia
;; COUNTRY: U.S.A.
;; ZIP: 20001
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/314,503A
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Posorske Esq., Laurence H.
;; REGISTRATION NUMBER: 34,698
;; REFERENCE/DOCKET NUMBER: 1107.47218
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202 508-9153
;; TELEFAX: 202 508-9299
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 759 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORIGINAL SOURCE:
;; ORGANISM: Mus sp
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..548
US-08-314-503A-4

Query Match 51.4%; Score 19; DB 1; Length 759;
Best Local Similarity 81.5%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatatccaccatccctccctc 27
||| | ||||| ||||| |||||
Db 591 CTGTTTACCCTATCCCTCCCT 617

RESULT 13
US-08-468-066-4
;; Sequence 4, Application US/08468066
;; Patent No. 5756676
;; GENERAL INFORMATION:
;; APPLICANT: Pasternack, Gary R.
;; APPLICANT: Kuhajda, Francis P.
;; TITLE OF INVENTION: No. 5756676el Mammalian Protein Associated With
;; TITLE OF INVENTION: Uncontrolled Cell Division
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner, Birch, McKie & Beckett
;; STREET: 1001 G Street, N.W.
;; CITY: Washington, D.C.
;; STATE: District of Columbia
;; COUNTRY: U.S.A.
;; ZIP: 20001
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/468,066
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/314,503
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Posorske Esq., Laurence H.
;; REGISTRATION NUMBER: 34,698
;; REFERENCE/DOCKET NUMBER: 1107.47218
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202 508-9153
;; TELEFAX: 202 508-9299
;; INFORMATION FOR SEQ ID NO: 4:

US-08-466-717-4

Qy 1 ctatatccaccatatacccccttccccct 27
||| | ||||| ||||| |||||
Db 591 CTGTTTATACCATATCCCCCTCCCCCT 617

Search completed: June 4, 2002, 17:47:29
Job time: 6602 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:56:22 ; Search time 521.5 Seconds
(without alignments)
121.814 Million cell updates/sec

Title: US-09-721-543A-16

Perfect score: 37

Sequence: 1 ctatatccaccatattcccttcccttgcgtgtggy 37

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 22.2 | 60.0 | 1302 | 23 | DNA encoding novel |
| 2 | 22.2 | 60.0 | 1302 | 23 | DNA encoding novel |
| 3 | 22.2 | 60.0 | 6337 | 23 | DNA encoding novel |
| 4 | 21 | 56.8 | 3132 | 22 | Human cDNA sequenc |
| c 5 | 20.8 | 56.2 | 619 | 21 | Fusarium venenatum |
| c 6 | 20.8 | 56.2 | 1462 | 22 | Human polynucleoti |
| 7 | 20.6 | 55.7 | 1452 | 23 | DNA encoding novel |
| 8 | 20.6 | 55.7 | 1860 | 23 | DNA encoding novel |
| c 9 | 20.6 | 55.7 | 1927 | 23 | DNA encoding novel |

| | | | | | |
|------|------|------|-------|----|----------|
| c 10 | 20.6 | 55.7 | 2285 | 23 | ABL23366 |
| c 11 | 20.6 | 55.7 | 5909 | 23 | AAS69432 |
| c 12 | 20.6 | 6012 | 23 | 23 | AAS94024 |
| c 13 | 20.6 | 55.7 | 7903 | 23 | AAS77690 |
| c 14 | 20.4 | 55.1 | 13094 | 22 | AAK69540 |
| c 15 | 20.2 | 54.6 | 2647 | 16 | AAQ8483 |
| c 16 | 20.2 | 54.6 | 11494 | 24 | AAS18559 |
| c 17 | 20.2 | 54.6 | 11494 | 24 | AAS18560 |
| c 18 | 20 | 54.1 | 492 | 22 | AAH29461 |
| c 19 | 20 | 54.1 | 26370 | 23 | ABL07082 |
| c 20 | 19.8 | 53.5 | 455 | 22 | ABA70980 |
| c 21 | 19.8 | 53.5 | 455 | 22 | ABA37400 |
| c 22 | 19.8 | 53.5 | 455 | 22 | AAK13256 |
| c 23 | 19.8 | 53.5 | 455 | 22 | AAK45226 |
| c 24 | 19.8 | 53.5 | 455 | 22 | AAI21198 |
| c 25 | 19.8 | 53.5 | 455 | 22 | AAI51177 |
| c 26 | 19.8 | 53.5 | 512 | 22 | ABA73367 |
| c 27 | 19.8 | 53.5 | 512 | 22 | ABA38177 |
| c 28 | 19.8 | 53.5 | 512 | 22 | AAK20793 |
| c 29 | 19.8 | 53.5 | 512 | 22 | AAK46943 |
| c 30 | 19.8 | 53.5 | 512 | 22 | AAI25788 |
| c 31 | 19.8 | 53.5 | 512 | 22 | AAI5780 |
| c 32 | 19.8 | 53.5 | 523 | 22 | ABA08360 |
| c 33 | 19.8 | 53.5 | 555 | 22 | AAH33403 |
| c 34 | 19.8 | 53.5 | 600 | 22 | ABA58626 |
| c 35 | 19.8 | 53.5 | 600 | 22 | ABA21298 |
| c 36 | 19.8 | 53.5 | 600 | 22 | AAK08099 |
| c 37 | 19.8 | 53.5 | 600 | 22 | AAK31975 |
| c 38 | 19.8 | 53.5 | 600 | 22 | AAI15592 |
| c 39 | 19.8 | 53.5 | 600 | 22 | AAI35699 |
| c 40 | 19.8 | 53.5 | 840 | 23 | AAS86137 |
| c 41 | 19.8 | 53.5 | 1096 | 23 | AAS69756 |
| c 42 | 19.8 | 53.5 | 1125 | 22 | ABA11802 |
| c 43 | 19.8 | 53.5 | 1497 | 23 | AAS86138 |
| c 44 | 19.8 | 53.5 | 3423 | 23 | AAS86141 |
| c 45 | 19.8 | 53.5 | 3899 | 19 | AAV63192 |

ALIGNMENTS

RESULT 1

AAS69409
ID AAS69409 standard; cDNA; 1302 BP.

XX AC AAS69409;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #5213.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG05222.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 5213; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1302 BP; 414 A; 285 C; 304 G; 299 T; 0 other;

Query Match 60.0%; Score 22.2; DB 23; Length 1302;
 Best Local Similarity 77.1%; Pred. No. 17;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Qy 3 atataccaccatattcccttcccttcggtgtggg 37
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1171 atataccaccatattcccttcccttcggtgtggg 1205

RESULT 2
 AAS93976
 ID AAS93976 standard; cDNA; 1302 BP.
 XX
 AC AAS93976;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29780.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG29789.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 29780; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1302 BP; 414 A; 285 C; 304 G; 299 T; 0 other;

Query Match 60.0%; Score 22.2; DB 23; Length 1302;
 Best Local Similarity 77.1%; Pred. No. 17;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Qy 3 atataccaccatattcccttcccttcggtgtggg 37
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1171 atataccaccatattcccttcccttcggtgtggg 1205

RESULT 3
 AAS94029
 ID AAS94029 standard; cDNA; 6337 BP.
 XX
 AC AAS94029;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29833.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG29842.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess

PT biodiversity -
 PS Claim 1; SEQ ID NO 29833; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6337 BP; 1813 A; 1504 C; 1508 G; 1512 T; 0 other;
 Query Match 60.0%; Score 22.2; DB 23; Length 6337;
 Best Local Similarity 77.1%; Pred. No. 21;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 3 ataccaccatccctcccttcgtggtggg 37
 ||||| ||||| ||||| ||||| |||||
 Db 1643 ataccaccatcccttcgtggtggg 1677
 RESULT 4
 ID AAH15189 standard; cDNA; 3132 BP.
 XX
 AC AAH15189;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:13279.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 13279; 2537pp + CD ROM; English.
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 3132 BP; 601 A; 943 C; 769 G; 819 T; 0 other;
 Query Match 56.8%; Score 21; DB 22; Length 3132;
 Best Local Similarity 82.8%; Pred. No. 57;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 5 atccaccatcccttccttcgtggtg 33
 ||||| ||||| ||||| ||||| |||||
 Db 1354 atcccccattccaccattcgtggtg 1382
 RESULT 5
 ID AAF08347/C standard; cDNA; 619 BP.
 XX
 AC AAF08347;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:870.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 XX 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 XX

CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7903 BP; 2316 A; 1835 C; 1850 G; 1902 T; 0 other;

Query Match 55.7%; Score 20.6; DB 23; Length 7903;
Best Local Similarity 74.3%; Pred. No. 92;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 3 atataccaccatataccctcccttcgtggtggg 37
Db 5477 atataccaccatatacccttcctcttcttattcg 5511

RESULT 14

AAK69540/c
ID AAK69540 standard; DNA; 13094 BP.

XX
AC AAK69540;

XX
DT 06-NOV-2001 (first entry)

XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24352.

DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX
PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0245610.
PR 08-NOV-2000; 2000US-0245611.
PR 08-NOV-2000; 2000US-0245613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -

XX Disclosure; SEQ ID NO 24352; 3071bp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.

XX Sequence 13094 BP; 2675 A; 3841 C; 4019 G; 2559 T; 0 other;

Query Match 55.1%; Score 20.4; DB 22; Length 13094;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 8 caccatatacccttcccttgcgtgtg 37
DB 5298 CACCCCATCCACCCCTCGAGTGTG 5269

RESULT 15

AAQ86483/c
ID AAQ86483 standard; cDNA; 2647 BP.

XX AAQ86483;

XX 23-NOV-1995 (first entry)

DE Human transcription factor E2F-2 cDNA.

XX E2F-2; transcription; factor; regulation; cell cycle;

KW retinoblastoma; tumour; suppressor; ss.

XX Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 429..1683

FT /*tag= a

FT /product= E2F-2 transcription factor.

XX GB2282814-A.

XX 19-APR-1995.

XX 07-OCT-1994; 94GB-0020283.

XX 13-OCT-1993; 93US-0136119.

XX (MERI) MERCK & CO INC.

XX Heimbrook DC, Ivey-Hoyle M, Oliff AI;

XX WPI; 1995-141220/19.

XX P-PSDB; AAR71734.

XX New human transcription factor E2F-2 - involved in cell cycle
XX regulation and useful for drug screening, also related cDNA,
XX plasmids and transformed cells.

XX Claim 3; Page 30-31; 53pp; English.

XX AAQ86483 codes for the human transcription factor E2F-2. E2F-2 is
XX involved in cell cycle regulation. In particular binding of E2F to
XX the retinoblastoma gene product (pRb) causes down regulation of the
XX transcription of any genes containing the E2F binding site. E2F-2 is
XX useful in the study of cell cycle regulation especially in the study
XX of pRb and certain viral oncogenes and oncoproteins.

XX Sequence 2647 BP; 508 A; 861 C; 795 G; 483 T; 0 other;

Query Match 54.6%; Score 20.2; DB 16; Length 2647;

Best Local Similarity 75.8%; Pred. No. 1.1e+02;

Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 5 atcaccatatacccttcccttgcgtgtg 37

DB 767 ATCCACTCTGATGCACCTTCCCTTGGGGTGG 735

Search completed: June 4, 2002, 17:56:36

Job time: 7114 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:48:43 ; Search time 2137.99 Seconds
(without alignments)
362.154 Million cell updates/sec

Title: US-09-721-543A-16
Perfect score: 37
Sequence: 1 ctatatccaccatccccctccctgcgtggtggg 37

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | Score | Match | Length | DB | ID | Description |
|--------|-------|-------|-------|--------|----|----|-------------|
| NO. | | | | | | | |

| | | | | | | | |
|----|------|--------|--------|----------|--------------------|--------------------|--------------------|
| 1 | 24.4 | 65.9 | 180313 | 2 | AC098141 | AC098141 Rattus no | |
| 2 | 23.8 | 64.3 | 40205 | 9 | AF045450 | AF045450 Homo sapi | |
| 3 | 64.3 | 142742 | 9 | AF121782 | AF121782 Homo sapi | | |
| 4 | 23.8 | 64.3 | 165382 | 9 | AF064860 | AF064860 Homo sapi | |
| 5 | 23.8 | 64.3 | 203715 | 2 | AC018807 | AC018807 Homo sapi | |
| 6 | 23.8 | 64.3 | 340000 | 9 | HS21C080 | AL163280 Homo sapi | |
| 7 | 23.2 | 62.7 | 140000 | 9 | AP001791 | AP001791 Homo sapi | |
| c | 8 | 23.2 | 62.7 | 141605 | 2 | AC013732 | AC013732 Homo sapi |
| c | 9 | 62.7 | 148332 | 9 | AC079586 | AC079586 Homo sapi | |
| c | 10 | 23.2 | 62.7 | 162301 | 2 | AC011023 | AC011023 Homo sapi |
| 11 | 23.2 | 62.7 | 180446 | 2 | AP001597 | AP001597 Homo sapi | |
| 12 | 22.6 | 61.1 | 128481 | 2 | AC094327 | AC094327 Rattus no | |
| c | 13 | 61.1 | 163338 | 9 | AL162426 | AL162426 Human DNA | |
| c | 14 | 22.6 | 61.1 | 172588 | 2 | AC007430 | AC007430 Homo sapi |
| c | 15 | 22.6 | 61.1 | 236135 | 2 | AC093559 | AC093559 Rattus no |
| c | 16 | 22.4 | 60.5 | 101559 | 2 | AC095646 | AC095646 Rattus no |
| 17 | 22.4 | 60.5 | 170359 | 9 | AC018545 | AC018545 Homo sapi | |
| c | 18 | 22.4 | 60.5 | 198653 | 2 | AC091474 | AC091474 Mus muscu |
| 19 | 22.2 | 60.0 | 37959 | 2 | AC102828 | AC102828 Mus muscu | |
| 20 | 22.2 | 60.0 | 49804 | 2 | AC101896 | AC101896 Mus muscu | |
| 21 | 22.2 | 60.0 | 69900 | 2 | AC101433 | AC101433 Mus muscu | |
| 22 | 22.2 | 60.0 | 75693 | 2 | AC026569 | AC026569 Homo sapi | |
| 23 | 22.2 | 60.0 | 154062 | 9 | AC022394 | AC022394 Homo sapi | |
| c | 24 | 22.2 | 60.0 | 164946 | 9 | AP000553 | AP000553 Homo sapi |
| 25 | 22.2 | 60.0 | 166447 | 9 | AC018751 | AC018751 Homo sapi | |
| c | 26 | 22.2 | 60.0 | 169237 | 9 | AC009516 | AC009516 Homo sapi |
| 27 | 22.2 | 60.0 | 172989 | 9 | AC096736 | AC096736 Homo sapi | |
| 28 | 22.2 | 60.0 | 180144 | 2 | AC094163 | AC094163 Rattus no | |
| 29 | 21.8 | 58.9 | 1210 | 10 | RATCRTXN | L15011 Rattus norv | |
| c | 30 | 21.8 | 58.9 | 39984 | 9 | AP000546 | AP000546 Homo sapi |
| c | 31 | 21.8 | 58.9 | 69251 | 2 | AC105064 | AC105064 Mus muscu |
| c | 32 | 21.8 | 58.9 | 86519 | 9 | AP001860 | AP001860 Homo sapi |
| c | 33 | 21.8 | 58.9 | 110000 | 9 | HSY237C10_1 | Continuation (2 of |
| c | 34 | 21.8 | 58.9 | 126480 | 2 | AC040954 | AC040954 Mus muscu |
| c | 35 | 21.8 | 58.9 | 129240 | 9 | AC083826 | AC083826 Homo sapi |
| c | 36 | 21.8 | 58.9 | 147521 | 2 | AC013680 | AC013680 Homo sapi |
| 37 | 21.8 | 58.9 | 157606 | 2 | AC069172 | AC069172 Homo sapi | |
| 38 | 21.8 | 58.9 | 161973 | 9 | HUAC002038 | AC097080 Homo sapi | |
| 39 | 21.8 | 58.9 | 175878 | 2 | AC097080 | AC097080 Rattus no | |
| 40 | 21.8 | 58.9 | 187681 | 2 | AC080124 | AC080124 Homo sapi | |
| 41 | 21.8 | 58.9 | 188872 | 9 | AC020750 | AC020750 Homo sapi | |
| c | 42 | 21.8 | 58.9 | 206192 | 9 | AL133173 | AL133173 Human DNA |
| c | 43 | 21.6 | 58.4 | 4504 | 10 | MMRPL | Y07941 M.musculus |
| c | 44 | 21.6 | 58.4 | 122416 | 9 | AL451030 | AL451030 Human DNA |
| c | 45 | 21.6 | 58.4 | 146187 | 2 | AC025854 | AC025854 Homo sapi |

ALIGNMENTS

| | | | | | | |
|------------|--|---|-----------|-----|--------|-----------------|
| RESULT | 1 | AC098141 | 180313 bp | DNA | linear | HTG 21-DEC-2001 |
| LOCUS | AC098141 | Rattus norvegicus clone CH230-175N2, *** SEQUENCING IN PROGRESS | | | | |
| DEFINITION | AC098141 | ***, 64 unordered pieces. | | | | |
| ACCESSION | AC098141.2 | GI:17975707 | | | | |
| VERSION | AC098141.2 | GI:17975707 | | | | |
| KEYWORDS | HTGS_PHASE1 | | | | | |
| SOURCE | Norway rat. | | | | | |
| ORGANISM | Rattus norvegicus | | | | | |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | | |
| | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; | | | | | |
| | Rattus. | | | | | |
| REFERENCE | 1 (bases 1 to 180313) | | | | | |
| AUTHORS | Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., | | | | | |
| | Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbarta, J., | | | | | |
| | Benton, J., Blinage, K., Blankenburg, K., Bonnin, D., Bouck, J., | | | | | |
| | Bowie, S., Brivea, M., Brown, M., Brown, M., Bryant, N.P., Buhay, C., | | | | | |
| | Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., | | | | | |
| | Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., | | | | | |
| | Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., | | | | | |
| | Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., | | | | | |
| | Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., | | | | | |

Duggan-Rocha, S., Durbán, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollaway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogutu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Slsson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 180313)
 Worley, K.C.

Direct Submission
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Dec 21, 2001 this sequence version replaced gi:16327847.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GGM0
 Center clone name: CH230-175N2
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to findPhrapList

Consensus quality: 152722 bases at least Q40
 Consensus quality: 158678 bases at least Q30
 Consensus quality: 163556 bases at least Q20
 Estimated insert size: 155476; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 64 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved

* 1 11538: contig of 11538 bp in length
 * 11539 11638: gap of unknown length
 * 24157 24157: contig of 12519 bp in length
 * 24158 24257: gap of unknown length
 * 24258 29851: contig of 5594 bp in length
 * 29852 29951: gap of unknown length

29952 35087: contig of 5136 bp in length
 35088 35187: gap of unknown length
 35188 40280: contig of 5093 bp in length
 40281 40380: gap of unknown length
 40381 46334: contig of 5954 bp in length
 46335 51283: contig of unknown length
 51284 51383: contig of 4849 bp in length
 51384 55429: contig of 4046 bp in length
 55430 55555: contig of 4026 bp in length
 55556 59655: gap of unknown length
 59656 64002: contig of 4347 bp in length
 64003 67138: contig of 3036 bp in length
 67139 67238: gap of unknown length
 67239 71909: contig of 4670 bp in length
 71909 72008: gap of unknown length
 72009 76112: contig of 4004 bp in length
 76113 79851: contig of 3739 bp in length
 79852 82810: contig of 2859 bp in length
 82811 85784: contig of 2873 bp in length
 85785 89323: contig of 3440 bp in length
 89324 92464: contig of 3041 bp in length
 92465 92565: gap of unknown length
 92566 96068: contig of 3503 bp in length
 96069 98010: contig of 1843 bp in length
 98011 101382: contig of 3272 bp in length
 101383 104070: contig of 2588 bp in length
 104071 104170: gap of unknown length
 104171 106714: contig of 2544 bp in length
 106715 106815: gap of unknown length
 106816 109660: contig of 2845 bp in length
 109661 111944: gap of unknown length
 111945 112044: contig of 2185 bp in length
 112045 114902: contig of 2858 bp in length
 114903 115003: contig of 2033 bp in length
 115004 117135: gap of unknown length
 117136 118712: contig of 1577 bp in length
 118713 119911: contig of 1099 bp in length
 119912 122491: contig of 2480 bp in length
 122492 122591: gap of unknown length
 122592 124374: contig of 1783 bp in length
 124375 124475: contig of 2150 bp in length
 124476 126625: gap of unknown length
 126626 129207: contig of 2483 bp in length
 129208 131564: contig of 2257 bp in length
 131565 131664: gap of unknown length
 131665 133609: contig of 1944 bp in length
 133610 133708: gap of unknown length
 133709 135292: contig of 1584 bp in length
 135293 135837: gap of unknown length
 135838 136937: contig of 1445 bp in length
 136938 139081: contig of 2144 bp in length
 139082 139181: gap of unknown length
 139182 140313: contig of 1132 bp in length
 140314 142806: contig of 2393 bp in length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

```

* 142807 142906: gap of unknown length
* 142907 144999: contig of 2093 bp in length
* 145000 145099: gap of unknown length
* 145100 146796: contig of 1697 bp in length
* 146797 146896: gap of unknown length
* 146897 148298: contig of 1402 bp in length
* 148299 148398: gap of unknown length
* 148399 150041: contig of 1643 bp in length
* 150042 150141: gap of unknown length
* 150142 152251: contig of 2110 bp in length
* 152252 152351: gap of unknown length
* 152352 153739: contig of 1388 bp in length
* 153740 153839: gap of unknown length
* 153840 155484: contig of 1645 bp in length
* 155485 155584: gap of unknown length
* 155585 157145: contig of 1561 bp in length
* 157146 157245: gap of unknown length
* 157246 158257: contig of 1012 bp in length
* 158258 158357: gap of unknown length
* 158358 159337: contig of 1180 bp in length
* 159338 159637: gap of unknown length
* 159638 160906: contig of 1269 bp in length
* 160907 161006: gap of unknown length
* 161007 162058: contig of 1052 bp in length
* 162059 162158: gap of unknown length
* 162159 163545: contig of 1387 bp in length
* 163546 163645: gap of unknown length
* 163646 165068: contig of 1423 bp in length
* 165069 165168: gap of unknown length
* 165169 166211: contig of 1043 bp in length
* 166212 166311: gap of unknown length
* 166312 167660: contig of 1349 bp in length
* 167661 167760: gap of unknown length

Query Match 65.9%; Score 24.4; DB 2; Length 180313;
Best Local Similarity 82.4%; Pred. No. 29;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 tatcacccatccctccctccctgctgtggg 37
   ||| ||||| ||||| ||||| ||| |||
Db 22480 TATTCACCATACCCCTCCCTTCCTTCGTGAG 22513

RESULT 2
AF045450
LOCUS Homo sapiens chromosome 21q22.3 cosmid Q11M15, complete sequence.
DEFINITION AF045450
ACCESSION AF045450
VERSION AF045450.1 GI:2895783
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 40205)
AUTHORS Taudien,S. and Rosenthal,A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40205)
AUTHORS Taudien,S., Nordstiek,G., Dagand,E., Hildmann,T., Drescher,B.,
Weber,J., Rosenthal,A. and Yaspo,M.L.
Direct Submission
TITLE Submitted (29-JAN-1998) Genome Analysis, Institut for Molecular
JOURNAL Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
     Location/Qualifiers
         1..40205
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="21"
             /map="21q22.3"
             /clone="cosmid Q11M15"
             complement(293..586)
             /rpt_family="AluSc"
             /evidence=not_experimental

repeat_region
complement(643..564)
/rpt_family="MERVL"
/evidence=not_experimental
complement(1053..1176)
/note="GenScan, score = 5.19%, comment = Internal_exon 124
bp frame: 2 phase: 1"
/evidence=not_experimental
complement(1251..1859)
/rpt_family="MERVL"
/evidence=not_experimental
complement(2296..3220)
/rpt_family="HERV16"
/evidence=not_experimental
2818..2889
/note="Xpound exon prediction, score = 82% (0%)"
/evidence=not_experimental
complement(3221..3658)
/rpt_family="MLT1C"
/evidence=not_experimental
3359..3395
/note="Xpound exon prediction, score = 84% (0%)"
/evidence=not_experimental
complement(3874..4176)
/rpt_family="AluJb"
/evidence=not_experimental
complement(4330..4433)
/rpt_family="L2"
/evidence=not_experimental
4585..4638
/note="GRAIL, score = 95.00%, comment = excellent shadow"
/evidence=not_experimental
complement(5113..7899)
/rpt_family="L1PAL1"
/evidence=not_experimental
complement(6119..5598)
/note="GRAIL, score = 63.00%, comment = good"
/evidence=not_experimental
complement(7927..8045)
/rpt_family="FLAM_A"
/evidence=not_experimental
8048..8178
/rpt_family="L2"
/evidence=not_experimental
8771..8876
/note="MZEf, score = 67.4%"
/evidence=not_experimental
complement(8809..9187)
/rpt_family="MER1A"
/evidence=not_experimental
8837..8876
/note="GRAIL, score = 55.00%, comment = good shadow"
/evidence=not_experimental
complement(9113..9213)
/note="GRAIL, score = 55.00%, comment = good"
/evidence=not_experimental
complement(9190..9258)
/rpt_family="L1MC3"
/evidence=not_experimental
9259..9647
/rpt_family="L1MC3"
/evidence=not_experimental
complement(10138..10427)
/rpt_family="MLT2F"
/evidence=not_experimental
10662..10875
/rpt_family="MER58A"
/evidence=not_experimental
complement(11072..11129)
/note="Xpound exon prediction, score = 89% (0%)"
/evidence=not_experimental
complement(11072..11156)
/note="GRAIL, score = 86.00%, comment = excellent"
/evidence=not_experimental

```

```

repeat_region 11478..11597
/evidence-not_experimental
/rpt_family="L2"
exon /evidence-not_experimental
11570..11704
/note="GRAIL, score = 99.000%, comment = excellent shadow"
/evidence-not_experimental
exon 11570..11704
/note="M2EF, score = 66.1%"
/evidence-not_experimental
repeat_region complement(11687..11735)
/rpt_family="MIR"
/evidence-not_experimental
repeat_region complement(11774..12320)
/rpt_family="MLT1E"
/evidence-not_experimental
exon complement(11828..11949)
/note="GRAIL, score = 50.000%, comment = good"
/evidence-not_experimental
repeat_region 12623..12843
/rpt_family="AluSx"
/evidence-not_experimental
repeat_region 13102..13457
/rpt_family="THE1B"
/evidence-not_experimental
exon complement(13481..13519)
/note="GRAIL, score = 65.000%, comment = good"
/evidence-not_experimental
repeat_region 15065..15188
/rpt_family="HERVL"
/evidence-not_experimental
misc_feature 15130..15378
/note="GC score = 10.60 (249bp)"
/note="Region: GC content"
/evidence-not_experimental
repeat_region 15369..16268
/rpt_family="MERVL"
/evidence-not_experimental
exon complement(16623..16714)
/note="GRAIL, score = 72.000%, comment = good"
/evidence-not_experimental
exon 16661..16730
/note="M2EF, score = 93.5%"
/evidence-not_experimental
repeat_region 16880..17318
/rpt_family="MLT1C"
/evidence-not_experimental
repeat_region 17326..17436
/rpt_family="L1R16C"
/evidence-not_experimental
repeat_region complement(17540..17675)
/rpt_family="MIR"
/evidence-not_experimental
repeat_region complement(17805..17864)
/rpt_family="MLT1C"
/evidence-not_experimental
exon 18076..18098
/note="GRAIL, score = 45.000%, comment = marginal shadow"
/evidence-not_experimental
exon complement(18219..18349)
/note="GenScan, score = 2.99%, comment = Internal_exon 131
bp frame: 1 phase: 2"
/evidence-not_experimental
repeat_region complement(18717..19211)
/rpt_family="MLT1C"
/evidence-not_experimental
exon complement(19252..19399)
/note="GRAIL, score = 71.000%, comment = good"
/evidence-not_experimental
exon 19269..19350
/note="GRAIL, score = 81.000%, comment = excellent shadow"
/evidence-not_experimental
exon complement(19379..19399)
/note="Xpound exon prediction, score = 60% (0%)"

```

```

exon /evidence-not_experimental
complement(19501..19881)
/note="M2EF, score = 50.6%"
/evidence-not_experimental
repeat_region 19455..20020
/note="homology = 100.00%, score = 38, counts = 2"
/rpt_type=tandem
/rpt_unit=ttcttcgcacactctctgttgagaggtcgggttatcgc
/evidence-not_experimental
repeat_region 20223..20643
/rpt_family="MLT1E"
/evidence-not_experimental
repeat_region 20645..20910
/rpt_family="MLT1F"
/evidence-not_experimental
repeat_region 20961..21007
/rpt_family="MLT1F"
/evidence-not_experimental
repeat_region 21483..21781
/rpt_family="AluJb"
/evidence-not_experimental
complement(21504..21756)
/note="M2EF, score = 79.4%"
/evidence-not_experimental
complement(22325..22806)
/rpt_family="MLT1C"

Query Match 64.3%; Score 23.8; DB 9; Length 40205;
Best Local Similarity 80.0%; Pred. No. 60;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 atatccaccatattcccttccttcgtgtggg 37
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20137 ATGTCCTTCATATTCCTTCTCTTGAGTGGG 20171

RESULT 3
AF121782 142742 bp DNA linear PRI 02-FEB-1999
LOCUS Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.
DEFINITION AF121782
ACCESSION AF121782
VERSION AF121782.1 GI:4210991
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 142742)
AUTHORS Taudien,S., Dagand,E., Hildmann,T., Nordstiek,G., Drescher,B.,
Schattevoy,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and
Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..142742
/organism="Homo sapiens"
/db_xref="taxon.9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 206A10"
3..118
/rpt_family="MIR"
repeat_region /evidence-not_experimental
complement(140..415)
repeat_region /rpt_family="MLT2E"
/evidence-not_experimental
complement(668..881)
repeat_region /rpt_family="MER58A"
/evidence-not_experimental
complement(1078..1162)
exon /note="GRAIL, score = 86.000%, comment = excellent shadow"
/evidence-not_experimental

```

| | | | |
|---------------|--|---------------|--|
| repeat_region | 1114..1403 /rpt_family="MLT1I" /evidence=not_experimental | exon | 9290..9371 /note="GRAIL, score = 81.000%, comment = excellent shadow" /evidence=not_experimental |
| repeat_region | 1481..1603 /rpt_family="L2" /evidence=not_experimental | repeat_region | 9966..10041 /note="homology = 100.00%, score = 38, counts = 2" /rpt_type=tandem |
| exon | 1576..1710 /note="GRAIL, score = 99.000%, comment = excellent" /evidence=not_experimental | repeat_region | 10244..10664 /rpt_family="MLT1E" /evidence=not_experimental |
| exon | 1576..1710 /note="MZEUF, score = 66.1%" /evidence=not_experimental | repeat_region | 10665..11028 /rpt_family="MLT1F" /evidence=not_experimental |
| repeat_region | complement(1693..1741) /rpt_family="MIR" /evidence=not_experimental | repeat_region | 11318..11417 /rpt_family="L1MC/D" /evidence=not_experimental |
| repeat_region | complement(1780..2326) /rpt_family="MLT1E" /evidence=not_experimental | repeat_region | 11503..11818 /rpt_family="AluSg" /evidence=not_experimental |
| exon | complement(1834..1955) /note="GRAIL, score = 50.000%, comment = good shadow" /evidence=not_experimental | exon | 12902..13180 /rpt_family="AluSx" /evidence=not_experimental |
| repeat_region | 2629..2831 /rpt_family="AluSx" /evidence=not_experimental | repeat_region | complement(13057..13161) /note="MZEUF, score = 76.1%" /evidence=not_experimental |
| repeat_region | 3012..3122 /rpt_family="MLT1J" /evidence=not_experimental | repeat_region | complement(13226..13364) /rpt_family="L1MC2" /evidence=not_experimental |
| repeat_region | 3123..3478 /rpt_family="THE1B" /evidence=not_experimental | repeat_region | complement(13613..13908) /rpt_family="AluSx" /evidence=not_experimental |
| repeat_region | 3479..3618 /rpt_family="MLT1J" /evidence=not_experimental | exon | complement(14317..14614) /rpt_family="LTR37A" /evidence=not_experimental |
| exon | complement(3502..3540) /note="GRAIL, score = 65.000%, comment = good shadow" /evidence=not_experimental | repeat_region | complement(14962..15036) /note="MZEUF, score = 93%" /evidence=not_experimental |
| exon | 4674..4781 /note="GRAIL, score = 42.000%, comment = marginal" /evidence=not_experimental | repeat_region | 15255..15299 /rpt_family="L1PA6" /evidence=not_experimental |
| exon | 4712..4749 /note="Xpound exon prediction, score = 70% (0%)" /evidence=not_experimental | exon | complement(15469..15497) /note="Xpound exon prediction, score = 62% (0%)" /evidence=not_experimental |
| repeat_region | 5385..6104 /rpt_family="HERVL" /evidence=not_experimental | repeat_region | 16045..16768 /rpt_family="MER21A" /evidence=not_experimental |
| exon | 6682..6751 /note="MZEUF, score = 95.1%" /evidence=not_experimental | exon | complement(16555..16655) /note="GRAIL, score = 54.000%, comment = good" /evidence=not_experimental |
| exon | 6746..6751 /note="GenScan, score = 1.83%, comment = Initial_exon 6 bp frame: 1 phase: 0" /evidence=not_experimental | repeat_region | 16823..16850 /rpt_family="MER34" /evidence=not_experimental |
| repeat_region | 6901..7339 /rpt_family="MLT1C" /evidence=not_experimental | repeat_region | complement(16851..17208) /rpt_family="THE1A" /evidence=not_experimental |
| repeat_region | 7347..7457 /rpt_family="LTR16C" /evidence=not_experimental | repeat_region | 17209..17632 /rpt_family="MER34" /evidence=not_experimental |
| repeat_region | complement(7561..7696) /rpt_family="MIR" /evidence=not_experimental | repeat_region | complement(17633..17969) /rpt_family="AluSg1" /evidence=not_experimental |
| repeat_region | complement(7819..8220) /rpt_family="MLT1J" /evidence=not_experimental | repeat_region | complement(18004..18334) /rpt_family="LTR17" /evidence=not_experimental |
| exon | 8097..8119 /note="GRAIL, score = 45.000%, comment = marginal" /evidence=not_experimental | repeat_region | complement(18335..18632) /rpt_family="HERV17" /evidence=not_experimental |
| repeat_region | complement(8738..9232) /rpt_family="MLT1C" /evidence=not_experimental | | |
| exon | complement(9273..9420) /note="GRAIL, score = 71.000%, comment = good" /evidence=not_experimental | | |

```

/existence-not_experimental
18352..18420
/notes="GRAIL, score = 51.000%, comment = good shadow"
/existence-not_experimental
complement(18397..18584)
/notes="GRAIL, score = 66.000%, comment = good"
/existence-not_experimental
complement(18643..24647)
/rpt_family="HERV17"
/existence-not_experimental
complement(19228..19361)
/notes="M2EF, score = 92.8%"
/existence-not_experimental
20900..21135
/notes="M2EF, score = 51.4%"
/existence-not_experimental
21150..21275
/notes="Xpound exon prediction, score = 74% (0%)"
/existence-not_experimental

Query Match      64.3%; Score 23.8; DB 9; Length 142742;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 atataccaccatccccctcccttcgctgtggtgg 37
Db 10158 ATGTCCTTCATATTCCTCTCTTTGAGTGTGGG 10192

RESULT 4
AF064860 AF064860 165382 bp DNA linear PRI 02-JUN-1998
LOCUS Homo sapiens chromosome 21q22.3 PAC 70124, complete sequence.
DEFINITION AF064860
ACCESSION AF064860
VERSION AF064860.1 GI:3171153
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165382)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..165382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
complement(<1..77)
/notes="GenScan, score = 3.23%, comment = Single-exon_gene
279 bp frame: 2 phase: 0"
/existence-not_experimental
complement(43..366)
/rpt_family="Ltr16A"
/existence-not_experimental
complement(485..648)
/rpt_family="MER63B"
/rpt_family="Ltr16A"
/existence-not_experimental
complement(977..1626)
/rpt_family="HERV1"
/existence-not_experimental
complement(1708..2125)
/rpt_family="HERV1"
/existence-not_experimental
complement(2836..3210)
/rpt_family="MLT1A1"
/existence-not_experimental
2928..3064
/notes="GRAIL, score = 64.000%, comment = good"

```

```

/existence-not_experimental
complement(3456..3559)
/rpt_family="L2"
/existence-not_experimental
4328..4468
/rpt_family="MER86"
/existence-not_experimental
5138..5597
/rpt_family="L1M4"
/existence-not_experimental
complement(5641..5850)
/rpt_family="MIR"
/existence-not_experimental
complement(6246..6452)
/rpt_family="Ltr10A"
/existence-not_experimental
complement(6763..6923)
/rpt_family="MIR"
/existence-not_experimental
7437..7538
/rpt_family="MSTA"
/existence-not_experimental
7542..8239
/rpt_family="Ltr8"
/existence-not_experimental
8242..8546
/rpt_family="MSTA"
/existence-not_experimental
complement(8677..8727)
/rpt_family="L2"
/existence-not_experimental
9938..10445
/rpt_family="MLT1D"
/existence-not_experimental
10148..10273
/notes="GRAIL, score = 50.000%, comment = good"
/existence-not_experimental
10451..10538
/rpt_family="MIR"
/existence-not_experimental
complement(10682..11042)
/rpt_family="MER7A"
/existence-not_experimental
11815..11944
/notes="homology = 99.20%, score = 123, counts = 26"
/rpt_type=tandem
/rpt_unit=tcctc
/existence-not_experimental
complement(11944..12237)
/rpt_family="AluSc"
/existence-not_experimental
complement(11973..12013)
/notes="Xpound exon prediction, score = 83% (0%)"
/existence-not_experimental
complement(12139..12170)
/notes="Xpound exon prediction, score = 77% (0%)"
/existence-not_experimental
12770..14551
/rpt_family="L1M45"
/existence-not_experimental
13473..13620
/notes="GRAIL, score = 51.000%, comment = good"
/existence-not_experimental
14602..14648
/notes="M2EF, score = 91.1%"
/existence-not_experimental
complement(14681..14973)
/rpt_family="AluSg"
/existence-not_experimental
complement(15289..15406)
/notes="GRAIL, score = 58.000%, comment = good"
/existence-not_experimental
complement(15297..15346)

```

```
/note="Xpound exon prediction, score = 87% (0%)"
/evidence-not_experimental
15436..16003
/rpt_family="L1ME"
/evidence-not_experimental
complement(16008..16227)
/rpt_family="MLT1A1"
/evidence-not_experimental
complement(16248..16421)
/rpt_family="L1PB3"
/evidence-not_experimental
16259..16359
/note="MZEf, score = 72.4%"
/evidence-not_experimental
complement(16283..16406)
/note="MZEf, score = 61%"
/evidence-not_experimental
complement(16429..16576)
/rpt_family="MLT1A1"
/evidence-not_experimental
16577..17146
/rpt_family="L1M4"
/evidence-not_experimental
17153..17451
/rpt_family="AluSc"
/evidence-not_experimental
17453..18912
/rpt_family="L1M4"
/evidence-not_experimental
17486..17522
/note="MZEf, score = 70.3%"
/evidence-not_experimental
18775..18868
/note="MZEf, score = 62.5%"
/evidence-not_experimental
18921..18984
/rpt_family="L1MC/D"
/evidence-not_experimental
18990..19261
/rpt_family="AluJb"
/evidence-not_experimental
19262..19415
/rpt_family="L1MC/D"
/evidence-not_experimental
complement(19413..19788)
/rpt_family="L1PA13"
/evidence-not_experimental
19820..20234
/rpt_family="L1MC/D"
/evidence-not_experimental
20293..20572
/rpt_family="L1MC/D"
/evidence-not_experimental
complement(20623..20924)
/rpt_family="AluJb"
/evidence-not_experimental
20810..20978
/note="MZEf, score = 62.6%"
/evidence-not_experimental
complement(21141..21424)
/rpt_family="AluJo"
/evidence-not_experimental
22062..22191
/rpt_family="L1PB1"
/evidence-not_experimental
22432..22504
/note="GAIL, score = 59.000%, comment = good"
/evidence-not_experimental
22572..22795
/rpt_family="MLT2"
/evidence-not_experimental
complement(22810..23112)
/rpt_family="AluJo"
```

```
/evidence-not_experimental
23130..23282
/rpt_family="MLT2"
/evidence-not_experimental
23690..23750
/note="MZEf, score = 95.1%"
/evidence-not_experimental
23906..24295
/rpt_family="MLT1h"
/evidence-not_experimental
24048..24211
/note="GAIL, score = 52.000%, comment = good"

Query Match      64.3%; Score 23.8; DB 9; Length 165382;
Best Local Similarity 80.0%; Pred. No. 48;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 atatccaccatccccctcccttgctggtggg 37
   || ||| ||||| ||||| ||||| ||||| |||||
Db 156149 ATGCTCTCCATATCCCTTCTTTGACTGTGGG 156183

RESULT 5
AC018807
LOCUS      203715 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-648J12, WORKING DRAFT
AC018807
VERSION    AC018807.5 GI:8748944
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 203715)
            Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 203715)
            Waterston,R.H.
TITLE      Direct Submission
JOURNAL

Submitted (03-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 27, 2000 this sequence version replaced gi:7235308.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0648J12
----- Summary Statistics -----
Sequencing vector: M13; 69%
Sequencing vector: plasmid; 31%
Chemistry: Dye-primer ET; 69% of reads
Chemistry: Dye-terminator Big Dye; 31% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 191142 bases at least Q40
Consensus quality: 195733 bases at least Q30
Consensus quality: 197952 bases at least Q20
Insert size: 190000; agarose-gel
Insert size: 201315; sum-of-contigs
Quality coverage: 4.28 in Q20 bases; agarose-gel
Quality coverage: 4.08 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
```



```

repeat_region complement(13598..13892)
/nc="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 14013..14131
/nc="CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 14561..14598
/nc="TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 14758..15215
/nc="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
exon 15331..15460
/gene="SH3BGR"
number=6
3'UTR 15427..15460
/gene="SH3BGR"
repeat_region 16255..16471
/nc="AluSg/x"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(18013..18312)
/nc="AluDb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
exon 18739..19157
/gene="SH3BGR"

```

Query Match 64.3%; Score 23.8; DB 9; Length 340000;
 Best Local Similarity 80.0%; Pred. NO. 43;
 Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 3 atatccaccatccctccctccctggtggtgg 37
||||| ||||| ||||| ||||| ||||| |||||
Db 234299 ATGCTCCATATCCCTCTCTTGTGAGTGGG 234333

```

```

RESULT 7
AP001791
LOCUS AP001791 140000 bp DNA linear PRI 08-DEC-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-427G12,
complete sequence.
ACCESSION AP001791
VERSION AP001791.4 GI:17425233
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-427G12.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140000)
AUTHORS Fujiyama,A., Yada,T., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Hattori,M., Ishii,K., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Dec 7, 2001 this sequence version replaced gi:9927282.
FEATURES
Location/Qualifiers
1..140000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-427G12"

```

BASE COUNT 45897 a 26034 c 24406 g 43663 t
 ORIGIN

```

Query Match 62.7%; Score 23.2; DB 9; Length 140000;
Best Local Similarity 77.8%; Pred. NO. 83;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 tatatccaccatccctccctccctggtggtgg 37
||||| ||||| ||||| ||||| ||||| |||||
Db 86668 TATATTCACCAATATCCACTCTCTCTACTGTGGG 86703

```

```

RESULT 8
AC013732/c
LOCUS AC013732 141605 bp DNA linear HTG 18-FEB-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-550023, WORKING DRAFT
SEQUENCE, 9 unordered pieces.
ACCESSION AC013732
VERSION AC013732.6 GI:9838265
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141605)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 141605)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:8961220.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0550023
----- Summary Statistics -----
Sequencing vector: M13; 73%
Sequencing vector: plasmid; 27%
Chemistry: Dye-primer ET; 73% of reads
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 136598 bases at least Q40
Consensus quality: 138339 bases at least Q30
Consensus quality: 139218 bases at least Q20
Insert size: 144000; agarose-fp
Insert size: 141800; sum-of-contigs
Quality coverage: 4.32 in Q20 bases; agarose-fp
Quality coverage: 4.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 5272: contig of 5272 bp in length
* 5273 5372: gap of unknown length
* 5373 9928: contig of 4556 bp in length
* 9929 10028: gap of unknown length
* 10029 19924: contig of 9896 bp in length
* 19925 20024: gap of unknown length
* 20025 33364: contig of 13340 bp in length
* 33365 33464: gap of unknown length
* 33465 47072: contig of 13608 bp in length
* 47073 47172: gap of unknown length
* 47173 67693: contig of 20521 bp in length

```



```

repeat_region      /rpt_family="L1"
1696. 1819
repeat_region      /rpt_family="MaLR"
1822. 2998
repeat_region      /rpt_family="L1"
2426. 2493
repeat_region      /rpt_family="(TA)n"
3013. 4163
repeat_region      /rpt_family="L1"
4179. 5284
repeat_region      /rpt_family="L1"
5278. 7922
repeat_region      /rpt_family="L1"
8761. 9068
repeat_region      /rpt_family="Alu"
9042. 9070
repeat_region      /rpt_family="(CAAA)n"
11123. 11170
repeat_region      /rpt_family="L2"
12097. 12675
repeat_region      /rpt_family="ERVL"
13615. 13669
repeat_region      /rpt_family="L2"
13715. 14327
repeat_region      /rpt_family="ERV1"
14667. 14906
repeat_region      /rpt_family="MaLR"
15849. 15968
repeat_region      /rpt_family="L2"
16293. 16601
repeat_region      /rpt_family="Alu"
17660. 17881
repeat_region      /rpt_family="ERVL"
18691. 18954
repeat_region      /rpt_family="L2"
19676. 19815
repeat_region      /rpt_family="MaLR"
19839. 19915
repeat_region      /rpt_family="L1"
19916. 20194
repeat_region      /rpt_family="Alu"
20195. 20543
repeat_region      /rpt_family="L1"
20344. 21003
repeat_region      /rpt_family="ERV1"
21004. 21189
repeat_region      /rpt_family="L1"
21302. 21609
repeat_region      /rpt_family="L1"
21577. 21729
repeat_region      /rpt_family="L1"
21730. 22027
repeat_region      /rpt_family="Alu"
22028. 23370
repeat_region      /rpt_family="L1"
23371. 23679
repeat_region      /rpt_family="Alu"
23680. 24606
repeat_region      /rpt_family="L1"
24660. 24928
repeat_region      /rpt_family="L1"
24840. 24867
repeat_region      /rpt_family="AT_rich"
24940. 25031
repeat_region      /rpt_family="(TA)n"
25342. 25646
repeat_region      /rpt_family="MaLR"
25725. 25766
repeat_region      /rpt_family="L1"
25767. 26172
repeat_region      /rpt_family="L1"
26154. 26174
repeat_region      /rpt_family="AT_rich"

```

```

repeat_region      27102. 27215
/rpt_family="MaLR"
28188. 28374
/rpt_family="MaLR"
29018. 29329
/rpt_family="Alu"
29315. 29341
/rpt_family="(TAAA)n"
29807. 29985
/rpt_family="L1"
30108. 30283
/rpt_family="L1"
30298. 30499
/rpt_family="L1"
30645. 30765
/rpt_family="L1"
30817. 31352
/rpt_family="L1"
31496. 31561
/rpt_family="L1"
31562. 31842
/rpt_family="Alu"
31839. 31862
/rpt_family="AT_rich"
31843. 32272
/rpt_family="L1"
32287. 32600
/rpt_family="L1"
32590. 32629
/rpt_family="(TTTTTC)n"

```

Query Match 62.7%; Score 23.2; DB 9; Length 148332;
 Best Local Similarity 77.8%; Pred. No. 82;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 cttatcccccattccctcccttcctgctgctgg 36
 | | | | | | | | | | | | | | | | | | | |
 Db 94047 CAACATCTACCCACATCCCTTCCTCCCTAGCCCTGG 94012

RESULT 10
 AC011023/c

LOCUS AC011023 Homo sapiens clone RP11-115C16, WORKING DRAFT SEQUENCE, 8 unordered
 DEFINITION HTG; HTGS_PHASE1; HTGS_DRAFT.

ACCESSION AC011023

VERSION AC011023.3 GI:7341743

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 162301)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 162301)

Birren,B., Linton,L., Nusbaum,C., Kann,L., Karatas,A., Klein,J.,

McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,N.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome

COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 30, 2000 this sequence version replaced gi:6573937.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1985
 Center clone name: L15.C16
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 157048 bases at least Q40
 Consensus quality: 159564 bases at least Q30
 Consensus quality: 160513 bases at least Q20
 Insert size: 194000; agarose-fp
 Insert size: 161601; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 5.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 5072: contig of 5072 bp in length
 * 5073 5172: gap of 100 bp
 * 5173 12256: contig of 7084 bp in length
 * 12257 12356: gap of 100 bp
 * 12357 20176: contig of 7820 bp in length
 * 20177 20276: gap of 100 bp
 * 20277 34761: contig of 14485 bp in length
 * 34762 34861: gap of 100 bp
 * 34862 63485: contig of 28624 bp in length
 * 63486 63585: gap of 100 bp
 * 63586 94394: contig of 30809 bp in length
 * 94395 94494: gap of 100 bp
 * 94495 126560: contig of 32066 bp in length
 * 126561 126660: gap of 100 bp
 * 126661 162301: contig of 35641 bp in length.
 * Location/Qualifiers
 1..162301
 /organism="Homo.sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-115C16"
 /clone.lib="RPC1-11 Human Male BAC"
 misc_feature 1..5072
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 5173..12256
 /note="assembly_fragment"
 12357..20176
 /note="assembly_fragment"
 20277..34761
 /note="assembly_fragment"
 34862..63485
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 63586..94394
 /note="assembly_fragment"
 94495..126560
 /note="assembly_fragment"
 126661..162301

/note="assembly_fragment"
 BASE COUNT 52447 a 29329 c 29494 g 50328 t 703 others
 ORIGIN
 Query Match 62.7%; Score 23.2; DB 2; Length 162301;
 Best Local Similarity 77.8%; Pred. No. 81;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 2 tatattccaccatattccctcccttgggtgtggg 37
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 112670 TATATTCCACCATATCCACTTCTCTCTACTGTGGG 112635
 RESULT 11
 LOCUS AP001997 180446 bp DNA linear HTG 30-MAY-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-281D19 map 11q, WORKING DRAFT
 SEQUENCE, 15 unordered pieces.
 ACCESSION AP001997
 VERSION AP001997.2 GI:8117612
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens DNA, clone:RP11-281D19.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 180446)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Vada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Homo sapiens 180,446 genomic DNA of 11q
 JOURNAL Published Only in Database (2000) In press
 REFERENCE 2 (bases 1 to 180446)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Vada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail: hattori@gscl.riken.go.jp,
 URL: <http://hgp.gsc.riken.go.jp/>, Tel:81-42-778-9923,
 Fax:81-42-778-9924)
 COMMENT On May 30, 2000 this sequence version replaced gi:7768922.
 ----- Genomic Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: <http://hgp.gsc.riken.go.jp/>
 Contact: hattori@gscl.riken.go.jp
 ----- Project Information
 Center project name: HumDrafl1
 Center clone name: RP11-281D19
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 170936 bases at least Q40
 Consensus quality: 175280 bases at least Q30
 Consensus quality: 177565 bases at least Q20
 Insert size: 179046; sum-of-contigs
 Quality coverage: 5.43x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of
 15 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved
 1 31405 contig of 31405 bp in length
 31506 54312 contig of 22807 bp in length
 54413 73827 contig of 19415 bp in length
 73928 91654 contig of 17727 bp in length
 91755 106526 contig of 14772 bp in length
 106627 121878 contig of 15252 bp in length

121979 132642 contig of 10664 bp in length
 132743 142911 contig of 10169 bp in length
 143012 152327 contig of 9316 bp in length
 152428 159844 contig of 7417 bp in length
 159945 165718 contig of 5774 bp in length
 165819 172001 contig of 6183 bp in length
 172102 175943 contig of 3842 bp in length
 176044 179290 contig of 3247 bp in length
 179391 180446 contig of 1056 bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 31405: contig of 31405 bp in length
 31406 31505: gap of 100 bp
 31506 54312: contig of 22807 bp in length
 54313 54412: gap of 100 bp
 54413 73827: contig of 19415 bp in length
 73828 73927: gap of 100 bp
 73928 91654: contig of 17727 bp in length
 91655 91754: gap of 100 bp
 91755 106526: contig of 14772 bp in length
 106527 106626: gap of 100 bp
 106627 121878: contig of 15252 bp in length
 121879 121978: gap of 100 bp
 121979 132642: contig of 10664 bp in length
 132643 132742: gap of 100 bp
 132743 142911: contig of 10169 bp in length
 142912 143011: gap of 100 bp
 143012 152327: contig of 9316 bp in length
 152328 152427: gap of 100 bp
 152428 159844: contig of 7417 bp in length
 159845 159944: gap of 100 bp
 159945 165718: contig of 5774 bp in length
 165719 165818: gap of 100 bp
 165819 172001: contig of 6183 bp in length
 172002 172101: gap of 100 bp
 172102 175943: contig of 3842 bp in length
 175944 176043: gap of 100 bp
 176044 179290: contig of 3247 bp in length
 179291 179390: gap of 100 bp
 179391 180446: contig of 1056 bp in length.

FEATURES

source
 1. .180446
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-281D19"
 1. .31405
 /note="assembly_fragment"
 31506. .54312
 /note="assembly_fragment clone_end:sp6 vector_side:left"
 54413. .73827
 /note="assembly_fragment"
 73928. .91654
 /note="assembly_fragment"
 91755. .106526
 /note="assembly_fragment"
 106627. .121878
 /note="assembly_fragment"
 121979. .132642
 /note="assembly_fragment"
 132743. .142911
 /note="assembly_fragment"
 143012. .152327
 /note="assembly_fragment"
 152428. .159844

misc_feature /note="assembly_fragment"
 159945. .165718
 /note="assembly_fragment"
 165819. .172001
 /note="assembly_fragment"
 172102. .175943
 /note="assembly_fragment clone_end:T7 vector_side:left"
 176044. .179290
 /note="assembly_fragment"
 179391. .180446
 /note="assembly_fragment"
 BASE COUNT 58765 a 32357 c 31761 g 56162 t 1401 others
 ORIGIN

Query Match 62.7%; Score 23.2; DB 2; Length 180446;
 Best Local Similarity 77.8%; Pred. No. 79;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 2 tatatccaccatcccccttcccccttgcgtgtggg 37
 ||||| ||||| ||||| ||||| || |||||
 Db 17658 TATATTCACCAATATCCACTTCCTCTACTGTGGG 17693

RESULT 12
 AC094327 128481 bp DNA linear HTG 20-DEC-2001
 LOCUS Rattus norvegicus clone CH230-3123, *** SEQUENCING IN PROGRESS ***,
 DEFINITION 67 unordered pieces.
 AC094327
 VERSION AC094327.2 GI:17941045
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
 1 (bases 1 to 128481)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homs,E., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
 Slisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Direct Submission
 Unpublished
 2 (bases 1 to 128481)
 Worley, K.C.

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15624160.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GAJW
 Center clone name: CH230-3123

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to
 findPhrapList

Consensus quality: 108264 bases at least Q40

Consensus quality: 116155 bases at least Q30

Consensus quality: 122636 bases at least Q20

Estimated insert size: 936774; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 67 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5834: contig of 5834 bp in length

5835 5934: gap of unknown length

5935 9357: contig of 3423 bp in length

9358 9457: gap of unknown length

9458 12732: contig of 3275 bp in length

12733 12832: gap of unknown length

12833 17592: contig of 4860 bp in length

17593 17792: gap of unknown length

17793 20226: contig of 2334 bp in length

20227 20226: gap of unknown length

20227 23758: contig of 3532 bp in length

23759 23858: gap of unknown length

23859 26574: contig of 2716 bp in length

26575 26674: gap of unknown length

26675 28571: contig of 1897 bp in length

28572 28572: gap of unknown length

28572 30715: contig of 2044 bp in length

30716 30815: gap of unknown length

30816 33288: contig of 2473 bp in length

33289 33388: gap of unknown length

33389 35335: contig of 2047 bp in length

35336 35535: gap of unknown length

35536 37858: contig of 2323 bp in length

37859 37958: gap of unknown length

37959 40008: contig of 2050 bp in length

40009 40108: gap of unknown length

40109 42562: contig of 2454 bp in length

42563 42662: gap of unknown length

42663 44720: contig of 2058 bp in length

44721 44820: gap of unknown length

44821 46542: contig of 1722 bp in length

46543 46642: gap of unknown length

46642 46642: gap of unknown length

46642 46642: gap of unknown length

46642 46642: gap of unknown length

46642 46642: gap of unknown length

46642 46642: gap of unknown length

46642 46642: gap of unknown length

46642 46642: gap of unknown length

48842: contig of 2200 bp in length
 48842: gap of unknown length
 50541: contig of 1599 bp in length
 50541: gap of unknown length
 50641: contig of 1360 bp in length
 52101: gap of unknown length
 52101: contig of 2693 bp in length
 54795: gap of unknown length
 54795: contig of 2143 bp in length
 57037: gap of unknown length
 57037: contig of 2180 bp in length
 59117: gap of unknown length
 59117: contig of 1192 bp in length
 60609: gap of unknown length
 60609: contig of 1254 bp in length
 61963: gap of unknown length
 61963: contig of 1202 bp in length
 63265: gap of unknown length
 63265: gap of unknown length
 64803: contig of 1438 bp in length
 64803: gap of unknown length
 66730: contig of 1827 bp in length
 66730: gap of unknown length
 68301: gap of unknown length
 68301: contig of 2110 bp in length
 69401: gap of unknown length
 69401: contig of 1208 bp in length
 70348: gap of unknown length
 70348: contig of 1882 bp in length
 72230: gap of unknown length
 72230: contig of 1750 bp in length
 74180: gap of unknown length
 74180: contig of 1176 bp in length
 75356: gap of unknown length
 75356: contig of 1379 bp in length
 76935: gap of unknown length
 76935: contig of 2174 bp in length
 79109: gap of unknown length
 79109: contig of 1552 bp in length
 80761: gap of unknown length
 80761: contig of 1006 bp in length
 81867: gap of unknown length
 81867: contig of 1577 bp in length
 83544: gap of unknown length
 83544: contig of 1623 bp in length
 85267: gap of unknown length
 85267: contig of 1155 bp in length
 86522: gap of unknown length
 86522: contig of 1756 bp in length
 88378: gap of unknown length
 88378: contig of 1208 bp in length
 89786: gap of unknown length
 89786: contig of 2125 bp in length
 91911: gap of unknown length
 91911: contig of 1793 bp in length
 93804: gap of unknown length
 93804: contig of 1703 bp in length
 95607: gap of unknown length
 95607: contig of 1271 bp in length
 96978: gap of unknown length
 96978: contig of 1746 bp in length
 98824: gap of unknown length
 98824: contig of 1417 bp in length
 100341: gap of unknown length
 100341: contig of 1166 bp in length
 101607: gap of unknown length
 101607: contig of 1503 bp in length
 103210: gap of unknown length
 103210: contig of 1598 bp in length
 104908: gap of unknown length
 104908: contig of 1083 bp in length
 106091: gap of unknown length
 106091: contig of 1386 bp in length
 107577: gap of unknown length
 107577: contig of 1023 bp in length
 108700: contig of 1023 bp in length


```
repeat_region 10849..11144
/note="AluSp repeat: matches 1..296 of consensus"
repeat_region 11145..11380
/note="L1ME3 repeat: matches 5884..6122 of consensus"
repeat_region 11565..11727
/note="FAM repeat: matches 1..163 of consensus"
repeat_region 11728..11805
/note="39 copies 2 mer ta 93% conserved"
repeat_region 11868..12182
/note="AluSc repeat: matches 1..309 of consensus"
repeat_region 12362..12677
/note="L2 repeat: matches 2342..2681 of consensus"
repeat_region 13235..13530
/note="AluSc repeat: matches 1..295 of consensus"
repeat_region 13531..13782
/note="AluSx repeat: matches 47..298 of consensus"
repeat_region 13799..13885
/note="AluSc repeat: matches 222..308 of consensus"
repeat_region 13944..14062
/note="7SLRNA repeat: matches 236..317 of consensus"
repeat_region 14066..14138
/note="L2 repeat: matches 2161..2240 of consensus"
repeat_region 14163..14310
/note="MIR repeat: matches 38..204 of consensus"
repeat_region 14349..14630
/note="AluJo repeat: matches 1..282 of consensus"
repeat_region 14671..14852
/note="L2 repeat: matches 2551..2743 of consensus"
repeat_region 14965..15097
/note="FLAM.C repeat: matches 1..133 of consensus"
repeat_region 15117..15325
/note="MSTD repeat: matches 193..394 of consensus"
repeat_region 15432..15483
/note="26 copies 2 mer tc 80% conserved"
repeat_region 15486..15659
/note="AluJo repeat: matches 123..299 of consensus"
repeat_region 15660..15968
/note="AluSg repeat: matches 1..309 of consensus"
repeat_region 15974..16005
/note="16 copies 2 mer tt 100% conserved"
repeat_region 16030..16167
/note="AluJo repeat: matches 1..138 of consensus"
repeat_region 16209..16519
/note="AluJo repeat: matches 3..310 of consensus"
repeat_region 16540..16923
/note="MER41D repeat: matches 1..435 of consensus"
repeat_region 16924..17238
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region 17239..17360
/note="MER41D repeat: matches 435..557 of consensus"
repeat_region 17533..17844
/note="AluSg repeat: matches 1..312 of consensus"
repeat_region 17872..18007
/note="AluJo repeat: matches 1..137 of consensus"
repeat_region 18008..18298
/note="AluSg repeat: matches 1..291 of consensus"
repeat_region 18299..18471
/note="AluJo repeat: matches 137..306 of consensus"
repeat_region 18605..18762
/note="AluSg/x repeat: matches 135..291 of consensus"
repeat_region 18763..19071
/note="AluSx repeat: matches 1..310 of consensus"
repeat_region 19072..19102
/note="AluSg/x repeat: matches 105..135 of consensus"
repeat_region 19104..19416
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 19551..19861
/note="AluJo repeat: matches 1..311 of consensus"
repeat_region 20477..20516
/note="MER4B repeat: matches 500..540 of consensus"
repeat_region 20541..20840
/note="AluSx repeat: matches 1..301 of consensus"
repeat_region 22154..22344
```

```
Query Match 61.1%; Score 22.6; DB 9; Length 163338;
Best Local Similarity 75.7%; Pred. No. 1.3e+02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

QY 1 ctatattccaccatattcccttcccttcggtgtggg 37

Db 21486 CHGTGTTACCCACATCCCTCTCTCTAGCCCTCTGGG 21450

RESULT 14

AC007430/c

LOCUS

DEFINITION

AC007430

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC007430 172588 bp DNA linear HTG 07-DEC-2001
Homo sapiens chromosome 9 clone RP11-94E8 map 9, *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces.

AC007430 G1:8705107

HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 172588)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 9, clone RP11-94E8

Unpublished

2 (bases 1 to 172588)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,

Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,

Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,

Donegan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,

Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,

Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,

Karatas,A., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P.,

Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,

Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,

Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,

Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,

Testaye,S., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,

Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission

Submitted (29-APR-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 25, 2000 this sequence version replaced gi:7342096.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L696

Center clone name: 94_E_8

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 2062 2161: contig of 2061 bp in length

* 2162 3140: gap of 100 bp

* 3141 3240: contig of 979 bp in length

* 3241 3240: gap of 100 bp

* 3241 6736: contig of 3496 bp in length

* 6737 6836: gap of 100 bp

* 6837 120051: contig of 113215 bp in length

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:09:02 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-15

Perfect score: 40

Sequence: 1 ctatttccaccatcccccttggtggcccttggtggtggg 40

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| c 1 | 24.2 | 60.5 | 288 | 9 | AI786367 |
| 2 | 24 | 60.0 | 528 | 9 | AU151810 |
| 3 | 23.4 | 58.5 | 754 | 9 | AA811417 |
| c 4 | 23 | 57.5 | 378 | 9 | BE142441 |
| 5 | 23 | 57.5 | 465 | 10 | BI014164 |
| 6 | 23 | 57.5 | 606 | 10 | BE531296 |
| 7 | 22.6 | 56.5 | 753 | 10 | BE774850 |
| c 8 | 22.6 | 56.5 | 929 | 10 | BM453604 |
| 9 | 22.4 | 56.0 | 281 | 10 | BI063933 |
| 10 | 22.4 | 56.0 | 1064 | 10 | BE260926 |
| c 11 | 22.2 | 55.5 | 369 | 10 | R25470 |
| 12 | 22.2 | 55.5 | 441 | 10 | BE490643 |
| 13 | 22.2 | 55.5 | 464 | 10 | BE444887 |
| c 14 | 22.2 | 55.5 | 489 | 10 | BE516924 |
| 15 | 22.2 | 55.5 | 616 | 10 | BE607929 |
| c 16 | 22.2 | 55.5 | 629 | 10 | BE585665 |
| c 17 | 22.2 | 55.5 | 1169 | 10 | BG341375 |

RESULT 1
AI786367/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

18 22 55.0 278 9 AI760326
19 22 55.0 442 9 AI207178
20 22 55.0 444 12 AQ625070
21 22 55.0 454 10 BF513006
c 22 22 55.0 464 10 T88687
23 22 55.0 465 9 AA973752
24 22 55.0 467 9 AA719824
25 22 55.0 519 9 BB771315
26 22 55.0 540 12 AQ116520
c 27 22 55.0 594 9 AL513845
c 28 22 55.0 599 9 AA469108
c 29 22 55.0 666 12 AG075780
c 30 22 55.0 791 9 AI765716
c 31 22 55.0 856 10 BG677396
c 32 21.8 54.5 349 12 PR0007209
c 33 21.8 54.5 460 9 AI139307
c 34 21.8 54.5 481 9 AI986176
c 35 21.8 54.5 895 9 AL522402
c 36 21.8 54.5 1039 10 BE962837
c 37 21.6 54.0 282 9 AW356590
c 38 21.6 54.0 385 9 AW428377
c 39 21.6 54.0 447 10 BG025725
c 40 21.6 54.0 467 10 R37656
c 41 21.6 54.0 512 10 BE757143
c 42 21.6 54.0 525 9 AA214934
c 43 21.6 54.0 545 12 AQ402382
c 44 21.6 54.0 553 9 AW656300
c 45 21.6 54.0 559 10 BE757518

ALIGNMENTS

AI786367 288 bp mRNA linear EST 02-JUL-1999
u54d05.x1 Sugano mouse liver mla Mus musculus cDNA clone
IMAGE:1923753 3' similar to gb:M73490 Mus musculus apolipoprotein E
mRNA, 3' end (MOUSE);, mRNA sequence.
AI786367
AI786367.1 GI:5334083
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 288)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School. of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:980045
Trace considered overall poor quality
Seq primer: custom primer used
High quality sequence stop: 1.
Location/Qualifiers
1. :288
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1923753"
/clone_lib="Sugano mouse liver mla"

DB 235 TCCACACGCCCTCGGCCCTTGGGTGGTGG 267

RESULT 4

BE142441/c 378 bp mRNA linear EST 21-JUN-2000
 LOCUS RC2-HT0148-071099-013-f01 HT0148 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE142441
 VERSION BE142441.1 GI:8605162
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 378)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC2-HT0148-071>)

099-013-f0143-1999-10-07&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 35
 High quality sequence stop: 378.
 Location/Qualifiers
 1..378

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0148"
 /dev_stage="Adult"

/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 107 a 83 c 93 g 95 t
 ORIGIN

Query Match 57.5%; Score 23; DB 9; Length 378;

Best Local Similarity 74.4%; Pred. No. 2.1e+02;

Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tatttccaccatattcccttggcccttgggtgtg 40

Db 43 TTTATCACACCAATGTCCTATGGTCCCTGGCTGTGGG 5

RESULT 5

BI014164
 LOCUS PM3-ET0207-090201-008-f04 ET0207 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BI014164
 VERSION BI014164.1 GI:14418235
 KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 465)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3-ET0207-090201-008-f04&t3=2001-02-09&t4=1>)

Seq primer: puc 18 forward
 High quality sequence start: 18
 High quality sequence stop: 465.
 Location/Qualifiers
 1..465

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0207"
 /dev_stage="Adult"

/note="Organ: lung tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 98 a 132 c 95 g 140 t
 ORIGIN

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0207"
 /dev_stage="Adult"

/note="Organ: lung tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 98 a 132 c 95 g 140 t
 ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Qy

Db

RESULT

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE531296
 601278222F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610492 5',
 mRNA sequence.
 BE531296
 BE531296.1 GI:9759855
 EST.
 human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 606)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.

Query Match 57.5%; Score 23; DB 10; Length 465;
 Best Local Similarity 74.4%; Pred. No. 2.2e+02;
 Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tatttccaccatattcccttggcccttgggtgtg 40
 Db 264 TTTATCACACCAATGTCCTATGGTCCCTGGCTGTGGG 302

RESULT 6
 BE531296
 LOCUS 601278222F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610492 5',
 mRNA sequence.
 BE531296
 BE531296.1 GI:9759855
 EST.
 human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 606)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.


```

RESULT 9
BI063933
LOCUS IL3-UT0117-060401-536-B07 UT0117 Homo sapiens cDNA, mRNA sequence.
DEFINITION IL3-UT0117-060401-536-B07 UT0117 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI063933
VERSION BI063933.1 GI:14471460
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 281)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalhal,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-UT0117-
060401-536-B07&t3=2001-04-06&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 281.
FEATURES
source
1..281
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT0117"
/dev_stage="Adult"
/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 91 a 53 c 53 g 84 t
ORIGIN
ctatttccaccatccatcccttggcccttgggtggg 40
||||| ||||||||| ||||| ||||| |||
3 CTTATCCCATTCATATCCCTTTGGACCTAGGAGAGG 42

RESULT 10
BE260926
LOCUS BE260926 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3509971 5',
DEFINITION 601153879P1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3509971 5',
mRNA sequence.
ACCESSION BE260926
VERSION BE260926.1 GI:9132626
KEYWORDS EST.
SOURCE human.

Query Match 56.0%; Score 22.4; DB 10; Length 281;
Best Local Similarity 72.5%; Pred. No. 3.4e-02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ctatttccaccatccatcccttggcccttgggtggg 40
||||| ||||||||| ||||| ||||| |||
3 CTTATCCCATTCATATCCCTTTGGACCTAGGAGAGG 42

Db 3 CTTATCCCATTCATATCCCTTTGGACCTAGGAGAGG 42

RESULT 11
R25470/c
LOCUS R25470/c
DEFINITION R25470
ACCESSION R25470
VERSION R25470.1 GI:781605
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 369)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1225
High quality sequence stops: 239

```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1064)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLC193 row: n column: 20

High quality sequence stop: 735.

Location/Qualifiers

1..1064

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3509971"

/clone_lib="NIH_MGC_19"

/tissue_type="neuroblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:

ECORI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC library."

230 a 345 c 277 g 212 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 81.2%; Score 22.4; DB 10; Length 1064;

Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 atttccaccatccatcccttggcccttggg 34

Db 885 ACTTCCACCAATATCCCTTTGGCCCGCGTG 916

RESULT 11

R25470/c

LOCUS

DEFINITION

R25470

IMAGE:132344 5', mRNA sequence.

ACCESSION

R25470

VERSION

R25470.1 GI:781605

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 369)

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman

,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J.,

Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston

,R., Williamson,A., Wohldmann,P. and Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1225

High quality sequence stops: 239

Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 Insert Length: 1225 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 239.
 Location/Qualifiers

FEATURES

1. .369
 /organism="Homo sapiens"
 /db_xref="GB:537936"
 /db_xref="taxon:9606"
 /clone="IMAGE:13234"
 /clone_lib="Soares placenta Nb2HP"
 /sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 AACTGGGAAGATTCCGCGCCGACGGAATTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 103 a 58 c 109 g 97 t 2 others
 ORIGIN
 Query Match 55.5%; Score 22.2; DB 10; Length 369;
 Best Local Similarity 77.1%; Pred. No. 4.2e+02;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ctattccaccatcccttggcccttgggt 35
 ||| ||||| ||| ||| ||||| ||||| |||||
 Db 306 CTACTTCCTTCCTTCCTCCTTGGTCTGGT 272

RESULT 12
 BE490643
 LOCUS
 DEFINITION WHE0370_B03_D062S Wheat cold-stressed seedling cDNA library
 Triticum aestivum cDNA clone WHE0370_B03_D06, mRNA sequence.

ACCESSION BE490643
 VERSION BE490643.1 GI:9610176
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.
 Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
 P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
 Seaton,C.L. and Tong,J.C.
 The structure and function of the expressed portion of the wheat
 genomes - Cold-stressed seedling cDNA library

JOURNAL Unpublished (2000)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818

Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.
 Location/Qualifiers

FEATURES

1. .441
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"

/clone="WHE0370_B03_D06"
 /clone_lib="Wheat cold-stressed seedling cDNA library"
 /tissue_type="Seedling"
 /dev_stage="Five-day old seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
 , germinated and grown aseptically in the dark at room
 temperature on filter paper with water, nystatin and
 cefotaxime in covered crystallization dishes. Five-day
 old seedlings were transferred to 5 C cold room and kept
 for 48 hr. The tissue, total RNA, and poly(A) RNA were
 prepared, a cDNA library was made, and the cDNA clones
 were in vivo excised to give phuescript phagemids in the
 TJ Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. Plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."

BASE COUNT 84 a 121 c 102 g 134 t
 ORIGIN

Query Match 55.5%; Score 22.2; DB 10; Length 441;
 Best Local Similarity 77.1%; Pred. No. 4.3e+02;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 4 ttccccaccatcccttggcccttgggtgtg 38
 ||| ||||| ||| ||| ||||| ||||| |||||
 Db 246 TTACCCCTTCCTTCCTTGGCTCTGGGTGTG 280

RESULT 13
 BE444887

LOCUS
 DEFINITION WHE1129_D06_G112S wheat etiolated seedling root normalized cDNA
 library Triticum aestivum cDNA clone WHE1129_D06_G11, mRNA
 sequence.

ACCESSION BE444887
 VERSION BE444887.1 GI:9444439
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.
 1 (bases 1 to 464)
 Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
 P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T.,
 Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.

The structure and function of the expressed portion of the wheat
 genomes - Normalized root cDNA library
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818

Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.
 Location/Qualifiers

1. .464
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1129_D06_G11"
 /clone_lib="Wheat etiolated seedling root normalized cDNA
 library"
 /tissue_type="Root"
 /dev_stage="Five day old etiolated seedling"
 /lab_host="E. coli DH10B"

FEATURES

1. .464
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1129_D06_G11"
 /clone_lib="Wheat etiolated seedling root normalized cDNA
 library"
 /tissue_type="Root"
 /dev_stage="Five day old etiolated seedling"
 /lab_host="E. coli DH10B"

Db 805 CTGTCACACATCTCTCTCTCAGCACGTCGGTGTGGG 844

RESULT 5

US-08-878-989-13

; Sequence 13, Application US/08878989

; Patent No. 5885803

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/878,989

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1866 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNUT03

; CLONE: 1340712

; US-08-878-989-13

Query Match 47.5%; Score 19; DB 2; Length 1866;

Best Local Similarity 71.4%; Pred. No. 50;

Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ttccaccaccatcccccttgggcccttggtggtg 39

| | | | | | | | | | | | | | | | | | | |

Db 165 TACCACCCACACCCCTTGCCCATTTTGGTCG 199

RESULT 6

US-09-272-796-13

; Sequence 13, Application US/09272796

; Patent No. 6207148

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/272,796

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/878,989

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1866 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNUT03

; CLONE: 1340712

; US-09-272-796-13

Query Match 47.5%; Score 19; DB 4; Length 1866;

Best Local Similarity 71.4%; Pred. No. 50;

Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ttccaccaccatcccccttgggcccttggtggtg 39

| | | | | | | | | | | | | | | | | | | |

Db 165 TACCACCCACACCCCTTGCCCATTTTGGTCG 199

RESULT 7

US-08-665-259-24/c

; Sequence 24, Application US/08665259

; Patent No. 6028173

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

```

, APPLICATION NUMBER: US 08/665,259
, FILING DATE: 17-JUN-1996
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: PCT/US96/10469
, FILING DATE: 17-JUN-1996
, ATTORNEY/AGENT INFORMATION:
, NAME: Dugan, Deborah A.
, REGISTRATION NUMBER: 37,315
, REFERENCE/DOCKET NUMBER: IG5-9.3
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (508) 872-8400
, TELEFAX: (508) 872-5415
, INFORMATION FOR SEQ ID NO: 24:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 5894 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, MOLECULE TYPE: CDNA
, FEATURE:
, NAME/KEY: CDS
, LOCATION: 2..5053
US-08-762-500-24

```

Query Match 47.5%; Score 19; DB 3; Length 5894;
Best Local Similarity 71.4%; Pred. No. 59;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 tatttccaccatatacccttgggccccttggtg 36
||||| | | ||||| | || |

RESULT 9
US-08-762-500-74/c
; Sequence 74, Application US/08762500

Sequence 74, Application US/08762500

APPLICANT: Connors, Timothy D.
 APPLICANT: Dackowski, William R.
 APPLICANT: Van Raay, Terence J.
 APPLICANT: Klinger, Katherine W.
 TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES.
 TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME.
 NUMBER OF SEQUENCES: 83
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENZYME CORPORATION
 STREET: One Mountain Road
 CITY: Framingham
 STATE: Massachusetts
 COUNTRY: United States of America
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/762,500
 FILING DATE: 09-DEC-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/665,259
 FILING DATE: 17-JUN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/10469
 FILING DATE: 17-JUN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Dugan, Deborah A.
 REGISTRATION NUMBER: 37,315

; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6525 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 573..5684
; US-08-762-500-74

Query Match 47.5%; Score 19; DB 3; Length 6525;

Best Local Similarity 71.4%; Pred. No. 60; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 10;

Qy 2 tatttccaccatccccctgggcccccttggtg 36
Db 2202 TATTCCCACTGAACACCTTGGACAGTGCTTG 2168

RESULT 10

US-08-991-789A-207/c
; Sequence 207, Application US/08991789A
; Patent No. 6225054

GENERAL INFORMATION:

; APPLICANT: Frudakis, Tony N.
; Smith, John M.
; Reed, Steven G.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 207:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 207:
; US-08-991-789A-207

Query Match 47.0%; Score 18.8; DB 4; Length 176;

Best Local Similarity 76.7%; Pred. No. 42; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 7;

Qy 1 ctatttccaccatcccccttgggcctt 30
Db 173 CCATTTACCACCCATATCCCATCAGTCACT 144

RESULT 11

US-09-062-451-207/c
; Sequence 207, Application US/09062451
; Patent No. 6344550

GENERAL INFORMATION:

; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 207:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-062-451-207

Query Match 47.0%; Score 18.8; DB 4; Length 176;

Best Local Similarity 76.7%; Pred. No. 42; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 7;

Qy 1 ctatttccaccatcccccttgggcctt 30
Db 173 CCATTTACCACCCATATCCCATCAGTCACT 144

RESULT 12

US-09-227-357-41/c
; Sequence 41, Application US/09227357
; Patent No. 6342581

GENERAL INFORMATION:

; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793

EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 41
LENGTH: 1158
TYPE: DNA
ORGANISM: Homo sapiens
US-09-227-357-41

Query Match 47.0%; Score 18.8; DB 4; Length 1158;
Best Local Similarity 76.7%; Pred. No. 56;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatttccaccatattcccttggccct 30
||| ||| ||| ||| ||| ||| ||| |||
Db 49 CTCTTCCCTCCCATTTCCCTCGTGCAT 20

RESULT 13
US-08-904-032-2/c
; Sequence 2, Application US/08904032
; Patent No. 6004773
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: EMBRYOGENESIS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/904,032
; APPLICATION NUMBER: US/08/904,032
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0359 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BLADNOT03
; CLONE: 1602555
US-08-904-032-2

Query Match 47.0%; Score 18.8; DB 3; Length 2921;
Best Local Similarity 76.7%; Pred. No. 64;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatttccaccatattcccttggccct 30
||| ||| ||| ||| ||| ||| ||| |||
Db 1882 CTGCTTCCACCCCAAGCCGCTTGGTTCT 1853

RESULT 14
US-09-134-246-8/c
; Sequence 8, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
; APPLICANT: Wayne, Jay
; APPLICANT: Xu, Shuang-yong
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; Vectors And Identification Of Two Thermus Plasmid

; TITLE OF INVENTION: Replication Origins
; FILE REFERENCE: Thermus Shuttle Vector
; CURRENT APPLICATION NUMBER: US/09/134,246B
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 11958
; TYPE: DNA
; ORGANISM: Thermus sp.
US-09-134-246-8

Query Match 47.0%; Score 18.8; DB 4; Length 11958;
Best Local Similarity 76.7%; Pred. No. 79;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctatttccaccatcccttggtggccct 30
|| ||||| || || ||||| |||||
Db 7859 CTTTTCCTCCCGCCTTTTCCTTGCCCT 7830

RESULT 15

US-09-123-851-2/c
; Sequence 2, Application US/09123851
; Patent No. 5958405
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL HUMAN PROTEIN TYROSINE KINASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/09/123,851
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/728,520
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0136 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1346 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-09-123-851-2

Query Match 46.0%; Score 18.4; DB 2; Length 1346;
Best Local Similarity 69.4%; Pred. No. 82;
Matches 25; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 ttccaccatcccttggtggcccttggtggtgg 39

Db 1108 TATGCCCCGCTTGCCCCCTGGGCCCTTGGGCTGG 1073
| | | | | | | | | | | | | | | | | | | | | |

Search completed: June 4, 2002, 17:47:26
Job time: 6599 sec

THIS PAGE BLANK (USPTO)

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>
RP5-1018K9 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-1018K9. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP5-1018K9 is at 1 in this sequence. The true left end of clone RP4-800F24 is at 15401 in this sequence.

FEATURES

| source | Location/Qualifiers |
|---------------|---|
| repeat_region | 1..125500 |
| repeat_region | /organism="Homo sapiens" |
| repeat_region | /db_xref="taxon:9606" |
| repeat_region | /chromosome="1" |
| repeat_region | /map="q23.2-24.3" |
| repeat_region | /clone="RP5-1018K9" |
| repeat_region | /clone_lib="RPCI-5" |
| repeat_region | 1..2074 |
| repeat_region | /note="L1MA4 repeat: matches 4221..6294 of consensus" |
| repeat_region | 2689..3162 |
| repeat_region | /note="L1PA15 repeat: matches 5683..6153 of consensus" |
| repeat_region | 3180..3610 |
| repeat_region | /note="L2 repeat: matches 2035..2495 of consensus" |
| repeat_region | 3677..3785 |
| repeat_region | /note="MER69A repeat: matches 1..110 of consensus" |
| repeat_region | 3792..3956 |
| repeat_region | /note="L2 repeat: matches 1740..1897 of consensus" |
| repeat_region | 3987..4199 |
| repeat_region | /note="L2 repeat: matches 2123..2335 of consensus" |
| repeat_region | 4403..4605 |
| repeat_region | /note="MIR repeat: matches 26..249 of consensus" |
| repeat_region | 4636..4734 |
| repeat_region | /note="L2 repeat: matches 2648..2749 of consensus" |
| repeat_region | 5440..5585 |
| repeat_region | /note="MIR repeat: matches 26..176 of consensus" |
| repeat_region | 5603..5798 |
| repeat_region | /note="MLT1F repeat: matches 352..539 of consensus" |
| repeat_region | 5919..6176 |
| repeat_region | /note="MLT1F repeat: matches 10..252 of consensus" |
| repeat_region | 6388..6637 |
| repeat_region | /note="MIR repeat: matches 3..262 of consensus" |
| repeat_region | 7575..7695 |
| repeat_region | /note="L2 repeat: matches 2623..2744 of consensus" |
| repeat_region | 10912..11284 |
| repeat_region | /note="MLT1A1 repeat: matches 20..365 of consensus" |
| repeat_region | 11990..12450 |
| repeat_region | /note="L2 repeat: matches 1..449 of consensus" |
| repeat_region | 12894..13515 |
| repeat_region | /note="match: STS: Em:G56281" |
| repeat_region | 12969..13077 |
| repeat_region | /note="L2 repeat: matches 2575..2690 of consensus" |
| repeat_region | 13538..13551 |
| repeat_region | /note="L2 repeat: matches 2256..2268 of consensus" |
| repeat_region | 13552..13868 |
| repeat_region | /note="AluJo repeat: matches 3..312 of consensus" |
| repeat_region | 13869..14262 |
| repeat_region | /note="L2 repeat: matches 2268..2679 of consensus" |
| repeat_region | 14286..14398 |
| repeat_region | /note="AluSg/x repeat: matches 168..294 of consensus" |
| repeat_region | 15207..15430 |
| repeat_region | /note="56 copies 4 mer cttt 79% conserved" |
| repeat_region | 15225..15432 |
| repeat_region | /note="104 copies 2 mer ct 77% conserved" |
| repeat_region | 15228..15419 |
| repeat_region | /note="6 copies 32 mer 83% conserved" |
| repeat_region | 15232..15434 |
| repeat_region | /note="7 copies 29 mer 66% conserved" |
| repeat_region | 15434..15669 |
| repeat_region | /note="AluJo repeat: matches 75..312 of consensus" |
| repeat_region | 15925..16102 |
| repeat_region | /note="MIR repeat: matches 75..260 of consensus" |
| repeat_region | 16637..16743 |
| repeat_region | /note="L2 repeat: matches 2586..2710 of consensus" |
| repeat_region | 16909..17295 |
| repeat_region | /note="LTR7 repeat: matches 1..405 of consensus" |
| repeat_region | 17661..18082 |
| repeat_region | /note="MSMB repeat: matches 1..426 of consensus" |
| repeat_region | 19420..19817 |
| repeat_region | /note="MLT1B repeat: matches 6..390 of consensus" |
| repeat_region | 21608..21667 |
| repeat_region | /note="20 copies 3 mer gag 78% conserved" |
| repeat_region | 22077..22227 |
| repeat_region | /note="MLT1B repeat: matches 359..504 of consensus" |
| repeat_region | 22323..22478 |
| repeat_region | /note="78 copies 2 mer aa 59% conserved" |
| repeat_region | 23525..24734 |
| repeat_region | /note="L1MD repeat: matches -15..1314 of consensus" |
| repeat_region | 24737..25041 |
| repeat_region | /note="AluJo repeat: matches 1..296 of consensus" |
| repeat_region | 25100..25570 |
| repeat_region | /note="MLT1D repeat: matches 5..503 of consensus" |
| repeat_region | 25807..25992 |
| repeat_region | /note="AluSg/x repeat: matches 118..304 of consensus" |
| repeat_region | 25993..26105 |
| repeat_region | /note="MLT1-INTERNAL repeat: matches 1019..1121 of consensus" |
| repeat_region | 26309..27011 |
| repeat_region | /note="L1M4 repeat: matches 3454..4224 of consensus" |
| repeat_region | 27012..27088 |
| repeat_region | /note="5S repeat: matches 1..89 of consensus" |
| repeat_region | 27094..27375 |
| repeat_region | /note="L1M4 repeat: matches 4216..4507 of consensus" |
| repeat_region | 27372..27878 |
| repeat_region | /note="L1M1 repeat: matches 4771..5297 of consensus" |
| repeat_region | 27891..28197 |
| repeat_region | /note="AluJb repeat: matches 13..312 of consensus" |
| repeat_region | 28219..28983 |
| repeat_region | /note="MER21B repeat: matches 71..788 of consensus" |
| repeat_region | 29017..29125 |
| repeat_region | /note="L1MD1 repeat: matches 6116..6224 of consensus" |
| repeat_region | 29805..29994 |
| repeat_region | /note="MIR repeat: matches 59..251 of consensus" |
| repeat_region | 30615..30873 |
| repeat_region | /note="MIR repeat: matches 13..261 of consensus" |
| repeat_region | 30916..30981 |
| repeat_region | /note="22 copies 3 mer tgg 84% conserved" |
| repeat_region | 31247..31537 |
| repeat_region | /note="AluSx repeat: matches 23..301 of consensus" |
| repeat_region | 31674..31970 |
| repeat_region | /note="AluSc repeat: matches 1..298 of consensus" |
| repeat_region | 32240..32562 |
| repeat_region | /note="AluSx repeat: matches 1..312 of consensus" |
| repeat_region | 33414..33457 |
| repeat_region | /note="11 copies 4 mer acac 88% conserved" |
| repeat_region | 34188..34484 |
| repeat_region | /note="AluSx repeat: matches 1..303 of consensus" |
| repeat_region | 34490..34614 |
| repeat_region | /note="L1M1 repeat: matches 5433..5569 of consensus" |
| repeat_region | 34616..34913 |
| repeat_region | /note="AluSx repeat: matches 13..311 of consensus" |

| | | |
|---------------|---------------|---|
| repeat_region | 34914. .35125 | /note="AluSg/x repeat: matches 97. .302 of consensus" |
| repeat_region | 35130. .35542 | /note="L1ME1 repeat: matches 5601. .6062 of consensus" |
| misc_feature | 35620. .36011 | /note="match: STS: Em.HSPE7ID11" |
| repeat_region | 35628. .35639 | /note="L1MB8 repeat: matches 6021. .6032 of consensus" |
| repeat_region | 35888. .36031 | /note="L1ME1 repeat: matches 5615. .5755 of consensus" |
| repeat_region | 36837. .36963 | /note="MCR57-Internal repeat: matches 7145. .7268 of consensus" |
| repeat_region | 37578. .37683 | /note="MIR repeat: matches 30. .145 of consensus" |
| repeat_region | 37725. .37849 | /note="MIR repeat: matches 20. .144 of consensus" |
| misc_feature | | complement(38957. .39396) |
| repeat_region | 39197. .39297 | /note="match: STS: Em.G07786" |
| repeat_region | 39303. .39329 | /note="FLAM.A repeat: matches 18. .115 of consensus" |
| repeat_region | 39490. .39533 | /note="9 copies 3 mer taa 96% conserved" |
| repeat_region | 39554. .39812 | /note="11 copies 4 mer ggtt 90% conserved" |
| repeat_region | 40316. .40341 | /note="AluJo repeat: matches 40. .297 of consensus" |
| repeat_region | 40785. .40853 | /note="13 copies 2 mer tt 92% conserved" |
| repeat_region | 40877. .40900 | /note="L1MD2 repeat: matches 5831. .5897 of consensus" |
| repeat_region | 40877. .40900 | /note="12 copies 2 mer tt 95% conserved" |

Query Match 57.5%; Score 23; DB 9; Length 125500;
Best Local Similarity 74.4%; Pred. No. 40;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tatttccacccatatacccttgggcccttggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | |
Db 81625 TTTATCACACCAATGTCCTATGTCCTCCCTGGTGTGGG 81663

| RESULT | 2 |
|------------|--|
| AL390855 | |
| LOCUS | 164736 bp DNA linear . HTG 13-JUN-2001 |
| DEFINITION | Homo sapiens chromosome 1 clone RP11-233O16, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces. |

AL390855.4 GI:9931008
 HTG; HTGS_PHASE1; HTGS_CANCELLED.
 human.

| SOURCE | ORGANISM | REFERENCE |
|--------|---|-----------|
| human. | Homo sapiens | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| | Mammali; Eutherii; Primates; Catarrhini; Hominidae; Homo. | |
| | 1 (bases 1 to 164736) | |

| | |
|---------|---|
| AUTHORS | FLUMSD, B. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk |
| COMMENT | On Aug 27, 2000 this sequence version replaced gi:9926766. |

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information

```

----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752;
Chemistry: Dye-terminator Big Dye; 1

```

Assembly program: XCAP4; version 4.3
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 162752 bases at least Q40
 Consensus quality: 163426 bases at least Q30
 Consensus quality: 163602 bases at least Q20
 Insert size: 164136; sum-of-contigs
 Insert size: 163581; 5.5% error; agarose-fp
 Quality coverage: 6.98x in Q20 bases; sum-of-contigs
 Quality coverage: 7.00x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be exact with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

| | | | |
|---|--------|---------------------------------------|-------------------------------|
| * | 1 | 50411: | contig. of 50411 bp in length |
| * | 50512 | 50511: gap of 100 bp | |
| * | 50512 | 57088: contig. of 6577 bp in length | |
| * | 57089 | 57188: gap of 100 bp | |
| * | 57189 | 89876: contig. of 32688 bp in length | |
| * | 89877 | 89976: gap of 100 bp | |
| * | 89977 | 95765: contig. of 5789 bp in length | |
| * | 95766 | 95865: gap of 100 bp | |
| * | 95866 | 116678: contig. of 20813 bp in length | |
| * | 116679 | 116778: gap of 100 bp | |
| * | 116779 | 151647: contig. of 34869 bp in length | |
| * | 151648 | 151747: gap of 100 bp | |
| * | 151748 | 164736: contig. of 12989 bp in length | |

FEATURES
source

```

1. 104750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-233016"
/clone_lib="RPC1-111"

```

misc_feature

misc_feature

```
misc_feature      50312. 37688  
                  /note="assembly_fragment:02233  
                  fragment_chain:1"  
misc_feature      57189. 89876
```

```
misc_feature
      071037. 850678
      /note="assembly_fragment:00092
      fragment_chain:1"
      89977. 95765
```

```
misc_feature 95866..116678
              /note="assembly_fragment:00719
              fragment_chain:1"
              95866..116678
```

```
misc_feature
/note="assembly_fragment:01774"
116779. 151647
/note="assembly_fragment:01730"
```

```
misc_feature
151748.164736
fragment_chain:2"
/note=assembly_fragment:01375
```

```
fragment_chain:2
clone_end:T7
vector_side:right"
```

| BASE COUNT | 50000 a | 32591 c | 31938 g | 49607 t | 600 others |
|------------|---------|---------|---------|---------|------------|
| ORIGIN | | | | | |

Query Match 57.5%; Score 23; DB 2; Length 164736;
Best Local Similarity 74.4%; Pred. No. 39;
Matches 29; Conservative 0; Mismatches 10; Indels 0

Qy 2 tatttcccccatatcccccttggccctt:gggtgtggg 40

Db 57345 TTTATCACACCAATGTCCCTATGGTCCCCCTGGCTGTGGG 57383

RESULT 3
AC098752/c
LOCUS
DEFINITION
AC098752
VERSION
KEYWORDS
SOURCE
ORGANISM

AC098752 65064 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-39G3, *** SEQUENCING IN PROGRESS ***
42 unordered pieces.
AC098752
AC098752.2 GI:17972917
HTG; HTGS_PHASE1.
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 65064)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Prantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseghe, H., Lozard, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogah, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Reyes, M., Rojas, A., Rojokan, I., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Slisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 65064)
Worley, K.C.
Direct Submission
Submitted (01-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16572781.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGT
Center clone name: CH230-39G3
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList

Consensus quality: 51113 bases at least Q40
Consensus quality: 54926 bases at least Q30
Consensus quality: 57793 bases at least Q20
Estimated insert size: 29709; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 42 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 1 2591: contig of 2591 bp in length
* 2592: gap of unknown length
* 2593: contig of 2811 bp in length
* 2594: gap of unknown length
* 5503: contig of 1730 bp in length
* 5504: gap of unknown length
* 7332: contig of 2706 bp in length
* 7333: gap of unknown length
* 10138: contig of 2706 bp in length
* 10139: gap of unknown length
* 12177: contig of 1939 bp in length
* 12178: gap of unknown length
* 12179: contig of 2167 bp in length
* 14444: gap of unknown length
* 14445: contig of 1108 bp in length
* 14545: gap of unknown length
* 15652: contig of 1108 bp in length
* 15653: gap of unknown length
* 15752: contig of 2127 bp in length
* 15753: gap of unknown length
* 17880: contig of 1509 bp in length
* 17881: gap of unknown length
* 19488: contig of 1509 bp in length
* 19489: gap of unknown length
* 19588: contig of 2165 bp in length
* 19589: gap of unknown length
* 21754: contig of 1042 bp in length
* 21854: gap of unknown length
* 22895: contig of 1275 bp in length
* 22896: gap of unknown length
* 24270: contig of 1275 bp in length
* 24271: gap of unknown length
* 24370: contig of 1161 bp in length
* 25531: gap of unknown length
* 25532: contig of 1881 bp in length
* 25631: gap of unknown length
* 27512: contig of 1447 bp in length
* 27612: gap of unknown length
* 27613: contig of 1447 bp in length
* 29059: gap of unknown length
* 29060: contig of 1049 bp in length
* 30208: gap of unknown length
* 30308: contig of 1334 bp in length
* 31642: gap of unknown length
* 31742: contig of 1101 bp in length
* 32843: gap of unknown length
* 32844: contig of 1041 bp in length
* 33984: gap of unknown length
* 34084: contig of 1837 bp in length
* 35921: gap of unknown length
* 36021: contig of 1179 bp in length
* 36022: gap of unknown length
* 37201: contig of 1092 bp in length
* 37301: gap of unknown length
* 38392: contig of 1453 bp in length
* 38492: gap of unknown length
* 39945: contig of 1017 bp in length
* 40045: gap of unknown length
* 41062: contig of 1017 bp in length
* 41063: gap of unknown length
* 41162: contig of 1036 bp in length
* 42198: gap of unknown length
* 42298: contig of 1125 bp in length
* 43423: gap of unknown length
* 43523: gap of unknown length
* 43524: contig of 1182 bp in length
* 44705: gap of unknown length
* 44803: gap of unknown length
* 45935: contig of 1130 bp in length
* 46035: gap of unknown length

```

* 46036 47316: contig of 1281 bp in length
* 47317 47416: gap of unknown length
* 47417 48893: contig of 1477 bp in length
* 48894 48993: gap of unknown length
* 48994 50069: contig of 1076 bp in length
* 50070 50169: gap of unknown length
* 50170 51313: contig of 1144 bp in length
* 51314 54133: gap of unknown length
* 54134 52416: contig of 1003 bp in length
* 52417 52516: gap of unknown length
* 52517 54086: contig of 1570 bp in length
* 54087 54186: gap of unknown length
* 54187 55690: contig of 1504 bp in length
* 55691 55790: gap of unknown length
* 55791 57121: contig of 1331 bp in length
* 57122 57221: gap of unknown length
* 57222 58288: contig of 1067 bp in length
* 58289 58388: gap of unknown length
* 58389 59001: contig of 1513 bp in length
* 59002 60001: gap of unknown length
* 60002 61061: contig of 1060 bp in length
* 61062 61161: gap of unknown length
* 61162 62494: contig of 1333 bp in length
* 62495 62594: gap of unknown length
* 62595 63732: contig of 1138 bp in length
* 63733 63832: gap of unknown length
* 63833 65064: contig of 1232 bp in length.

```

```

FEATURES
    source
        1..65064
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-3963"

```

```

BASE COUNT 16376 a 14619 c 14087 g 15808 t 4174 others
ORIGIN

```

```

Query Match 57.0%; Score 22.8; DB 2; Length 65064;
Best Local Similarity 79.4%; Pred. No. 50;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 4 ttctccaccatcatccctggccctgggtgt 37
|||||
DB 21611 TTTCACCCCATATCCCTTTGGCTGCGCCTCT 21578
|||||

```

```

RESULT 4
AL355994/c 157095 bp DNA linear HTG 10-AUG-2001
LOCUS Homo sapiens chromosome 1 clone RP11-25414, *** SEQUENCING IN
DEFINITION PROGRESS ***, 11 unordered pieces.
ACCESSION AL355994
VERSION AL355994.4 GI:10039855
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    Melay, K.
    Direct Submission
    Submitted (09-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
    requests: clonerequest@sanger.ac.uk
    On Sep 9, 2000 this sequence version replaced gi:9588227.
    ----- Genome Center
    Center: Sanger Centre
    Center code: SC
    Web site: http://www.sanger.ac.uk
    Contact: humquery@sanger.ac.uk
    ----- Project Information
    Center project name: BA25414
    ----- Summary Statistics
    Assembly program: XGAP4; version 4.5

```

```

Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 149592 bases at least Q40
Consensus quality: 152642 bases at least Q30
Consensus quality: 154560 bases at least Q20
Insert size: 156095; sum-of-contigs
Insert size: 156608; 3.9% error; agarose-fp
Quality coverage: 4.04x in Q20 bases; sum-of-contigs Quality
coverage: 4.13x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 2590: contig of 2590 bp in length
* 2591 2690: gap of 100 bp
* 2691 10620: contig of 7930 bp in length
* 10621 10720: gap of 100 bp
* 10721 39969: contig of 29249 bp in length
* 39970 40069: gap of 100 bp
* 40070 87137: contig of 47068 bp in length
* 87138 87237: gap of 100 bp
* 87238 99117: contig of 11880 bp in length
* 99118 99217: gap of 100 bp
* 99218 112196: contig of 12979 bp in length
* 112197 112296: gap of 100 bp
* 112297 116983: contig of 4687 bp in length
* 116984 117083: gap of 100 bp
* 117084 121151: contig of 4068 bp in length
* 121152 121251: gap of 100 bp
* 121252 134800: contig of 13549 bp in length
* 134801 134900: gap of 100 bp
* 134901 152651: contig of 17751 bp in length
* 152652 152751: gap of 100 bp
* 152752 157095: contig of 4344 bp in length.

```

FEATURES

source

```

1..157095
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-25414"
/clone_lib="RPC1-11.1"
1..2590
/note="assembly_fragment:00613
fragment_chain:1"
2691..10620
/note="assembly_fragment:00394
fragment_chain:1"
10721..39969
/note="assembly_fragment:01099
fragment_chain:1"
40070..87137
/note="assembly_fragment:00143
fragment_chain:1"
87238..99117
/note="assembly_fragment:00340
fragment_chain:2"
99218..112196
/note="assembly_fragment:00372
fragment_chain:2"
112297..116983
/note="assembly_fragment:00457
fragment_chain:3"
117084..121151
/note="assembly_fragment:01157
fragment_chain:3"
121252..134800
/note="assembly_fragment:00625
fragment_chain:4"

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature 134901..152651
/note="assembly_fragment:01354
fragment_chain:4"
misc_feature 152752..157095
/note="assembly_fragment:01275
clone_end:SP6
vector_side:right"
BASE COUNT 37323 a 41261 c 41264 g 35885 t 1002 others
ORIGIN

Query Match 57.08; Score 22.8; DB 2; Length 157095;
Best Local Similarity 79.4%; Pred. No. 47;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctatttccaccatcccttggcccttggg 34
|||||
Db 76307 CTAATCCACCCAGCTCCCTGAGGCCCTGGG 76274

RESULT 5
MMU304861
LOCUS
DEFINITION Mus musculus partial Bfsp2 gene for phakinin (CP49 protein), exon
1.
ACCESSION AJ304861.1 GI:17977855
VERSION Bfsp2 gene; CP49 protein; intermediate filament protein; phakinin.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Sandilands,A. and Quinlan,R.A.
TITLE Sequence analysis of the mouse CP49 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 24180)
AUTHORS Sandilands,A.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Sandilands A., Biochemistry, School of Life
Sciences, Dow Street, DD4 7AS, UNITED KINGDOM
FEATURES
source 1..24180
/organism="Mus musculus"
/strain="129"
/db_xref="taxon:10090"
7537..8028
/gene="Bfsp2"
7537..8028
/gene="Bfsp2"
/number=1
7537..8028
/gene="Bfsp2"
7537..8028
/gene="Bfsp2"
7537..8028
/gene="Bfsp2"
/function="intermediate filament protein"
/codon_start=1
/product="phakinin, CP49"
/protein_id="CAC83162.1"
/db_xref="GI:1797856"
/translation="MSKRRVAADLPSTNSMPVQRHVRVSSLRGTHSPSSLDSPFASR
TSAGSLVRAPGVYVAPSPGGGLGARVTRRALGISYVFLQGLRSSGLANYPAFCP
ERDHTTVDLGGCLVEYMTKVALEQVSOBLETLQRAHLESKAKSGGMDALRASWAS
SYQQ"
BASE COUNT 6533 a 5929 c 5535 g 6183 t
ORIGIN

Query Match 56.5%; Score 22.6; DB 10; Length 24180;
Best Local Similarity 75.7%; Pred. No. 64;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 ttctccaccatatacccttgggccccttgggtg 40
|||||
Db 4101 TGTCCCCCAACCCCTTGTTCCTGGTGGCGG 4137

RESULT 6
AC083820/c
LOCUS
DEFINITION Rattus norvegicus clone RP32-290H10, WORKING DRAFT SEQUENCE, 4
unordered pieces.
ACCESSION AC083820
VERSION 139572 bp DNA linear HTG 24-NOV-2001
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 139572)
AUTHORS Muzny,D.N., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Focha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,D.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louleghed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nkokenko,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmami,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 139572)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 24, 2001 this sequence version replaced gi:13346542.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUAIE

```


Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 2 (bases 1 to 67560)
 Worley, K.C.

JOURNAL REFERENCE AUTHORS

JOURNAL TITLE

Submitted (17-Sep-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15627285.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GCYX
 Center clone name: CH230-8P8
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329first call to
 findPhrapList

Consensus quality: 54982 bases at least Q40
 Consensus quality: 59759 bases at least Q30
 Consensus quality: 63487 bases at least Q20
 Estimated insert size: 48793; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 38 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2811: contig of 2811 bp in length
 2812 2911: gap of unknown length
 2912 5574: contig of 2663 bp in length
 5575 5674: gap of unknown length
 5675 6898: contig of 1224 bp in length
 6899 6998: gap of unknown length
 6999 10205: contig of 3207 bp in length
 10206 10306: gap of unknown length
 10306 12288: contig of 1983 bp in length
 12289 12389: gap of unknown length
 12389 15169: contig of 2781 bp in length
 15170 15269: gap of unknown length
 15270 17310: contig of 2041 bp in length
 17311 17410: gap of unknown length
 17411 19519: contig of 2108 bp in length
 19519 19619: gap of unknown length
 19619 21451: contig of 1832 bp in length
 21451 21551: gap of unknown length
 21551 24017: contig of 2466 bp in length
 24017 24117: gap of unknown length
 24117 26517: contig of 2400 bp in length
 26517 26617: gap of unknown length
 26617 27922: contig of 1305 bp in length
 27922 28022: gap of unknown length
 28022 29552: contig of 1530 bp in length
 29552 29652: gap of unknown length
 29652 30967: contig of 1315 bp in length
 30967 31067: gap of unknown length
 31067 32927: contig of 1860 bp in length
 32927 33027: gap of unknown length
 33027 34992: contig of 1965 bp in length
 34992 35091: gap of unknown length

* 35092 36254: contig of 1163 bp in length
 36255 36354: gap of unknown length
 36355 37841: contig of 1487 bp in length
 37842 37941: gap of unknown length
 37942 39780: contig of 1839 bp in length
 39781 39880: gap of unknown length
 39881 40913: contig of 1033 bp in length
 40914 41013: gap of unknown length
 41014 42258: contig of 1245 bp in length
 42259 42358: gap of unknown length
 42359 43288: contig of 1270 bp in length
 43289 43629: gap of unknown length
 43630 45210: contig of 1481 bp in length
 45211 45309: gap of unknown length
 45310 46726: contig of 1416 bp in length
 46727 48161: gap of unknown length
 48162 48261: gap of unknown length
 48262 49952: contig of 1691 bp in length
 49953 50053: gap of unknown length
 50054 51149: contig of 1097 bp in length
 51150 51249: gap of unknown length
 51250 52985: contig of 1736 bp in length
 52986 53086: gap of unknown length
 53087 55192: contig of 2107 bp in length
 55193 55293: gap of unknown length
 55294 57279: contig of 1987 bp in length
 57280 57380: gap of unknown length
 57381 58527: contig of 1147 bp in length
 58528 58626: gap of unknown length
 58627 59848: contig of 1222 bp in length
 59849 61138: gap of unknown length
 61139 61239: contig of 1190 bp in length
 61240 62287: contig of 1049 bp in length
 62288 62388: gap of unknown length
 62389 63441: contig of 1053 bp in length
 63442 63541: gap of unknown length
 63542 64591: contig of 1050 bp in length
 64592 64691: gap of unknown length
 64692 66335: contig of 1644 bp in length
 66336 66435: gap of unknown length
 66436 67560: contig of 1126 bp in length.

FEATURES source

Location/Qualifiers
 1..67560
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-8P8"
 BASE COUNT 19966 a 12041 c 11986 g 19717 t 3850 others
 ORIGIN

Query Match 56.0%; Score 22.4; DB 2; Length 67560;
 Best Local Similarity 72.5%; Pred. No. 73;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 1 ctatttccaccatattcccttgcccttggtgggtggg 40
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 56938 CTCCTTCCCAACCCCTACCCATTGGTCCCTGTGTGAGAAGG 56899

RESULT 10

AC102525/c LOCUS 70165 bp DNA linear HTG 23-NOV-2001
 DEFINITION MUS musculus clone RP24-133K20, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC102525
 VERSION AC102525.1 GI:17061611
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 70165)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-133K20
2 (bases 1 to 70165)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarato, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
MacLaren, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
Lamazzari, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18959
Center clone name: 133_K20

* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 671: contig of 671 bp in length
* 772 771: gap of 100 bp
* 772 1459: contig of 688 bp in length
* 1460 1559: gap of 100 bp
* 1560 2254: contig of 695 bp in length
* 2255 2354: gap of 100 bp
* 2355 3037: contig of 673 bp in length
* 3028 3127: gap of 100 bp
* 3128 3834: contig of 707 bp in length
* 3835 3934: gap of 100 bp
* 3935 4634: contig of 700 bp in length
* 4635 4734: gap of 100 bp
* 4735 5433: contig of 699 bp in length
* 5434 5533: gap of 100 bp
* 5534 6232: contig of 699 bp in length
* 6233 6332: gap of 100 bp
* 6333 7034: contig of 702 bp in length
* 7035 7134: gap of 100 bp
* 7135 7842: contig of 708 bp in length
* 7843 7942: gap of 100 bp
* 7943 8604: contig of 662 bp in length
* 8605 8704: gap of 100 bp
* 8705 9399: contig of 695 bp in length
* 9400 9499: gap of 100 bp
* 9500 10193: contig of 694 bp in length
* 10194 10993: gap of 100 bp
* 10994 10986: contig of 693 bp in length
* 10987 11086: gap of 100 bp
* 11087 11759: contig of 673 bp in length
* 11760 11859: gap of 100 bp
* 11860 12555: contig of 696 bp in length
* 12556 12655: gap of 100 bp
* 12656 13343: contig of 688 bp in length
* 13344 13443: gap of 100 bp
* 13444 14147: contig of 704 bp in length
* 14148 14247: gap of 100 bp
* 14248 14930: contig of 683 bp in length
* 14931 15030: gap of 100 bp
* 15031 15702: contig of 672 bp in length
* 15703 15802: gap of 100 bp
* 15803 16477: contig of 675 bp in length
* 16378 16577: gap of 100 bp
* 16378 17253: contig of 676 bp in length
* 17254 17353: gap of 100 bp
* 17354 18045: contig of 692 bp in length
* 18046 18145: gap of 100 bp
* 18146 18840: contig of 695 bp in length
* 18941 19616: contig of 676 bp in length
* 19617 19716: gap of 100 bp
* 19717 20423: contig of 707 bp in length
* 20424 20523: gap of 100 bp
* 20524 21200: contig of 677 bp in length
* 21201 21300: gap of 100 bp
* 21301 22001: contig of 701 bp in length
* 22002 22101: gap of 100 bp
* 22102 22799: contig of 698 bp in length
* 22800 22999: gap of 100 bp
* 22900 23597: contig of 698 bp in length
* 23598 23697: gap of 100 bp
* 23698 24396: contig of 699 bp in length
* 24397 24496: gap of 100 bp
* 24497 25167: contig of 671 bp in length
* 25168 25267: gap of 100 bp
* 25268 25956: contig of 689 bp in length
* 25957 26056: gap of 100 bp
* 26057 26742: contig of 686 bp in length
* 26743 26842: gap of 100 bp
* 26843 27534: contig of 692 bp in length
* 27535 27634: gap of 100 bp
* 27635 28323: contig of 689 bp in length
* 28324 28423: gap of 100 bp
* 28424 29134: contig of 711 bp in length
* 29135 29234: gap of 100 bp
* 29235 29882: contig of 648 bp in length
* 29883 29982: gap of 100 bp
* 29983 30672: contig of 690 bp in length
* 30673 30772: gap of 100 bp
* 30773 31476: contig of 704 bp in length
* 31477 31576: gap of 100 bp
* 31577 32258: contig of 682 bp in length
* 32259 32358: gap of 100 bp
* 32359 33040: contig of 682 bp in length
* 33041 33140: gap of 100 bp
* 33141 33816: contig of 676 bp in length
* 33817 33916: gap of 100 bp
* 33917 34584: contig of 668 bp in length
* 34585 34684: gap of 100 bp
* 34685 35360: contig of 676 bp in length
* 35361 35460: gap of 100 bp
* 35461 36136: contig of 676 bp in length
* 36137 36236: gap of 100 bp
* 36237 36932: contig of 696 bp in length
* 36933 37032: gap of 100 bp

TITLE Homo sapiens genomic DNA, chromosome 21q
JOURNAL Published Only in DataBase (1998) In press
REFERENCE 2 (bases 1 to 100000)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) to the DBJ/EMBL/GenBank databases.
JOURNAL Masahira Hattori, Kitasato University, Department of Science, JST
JOURNAL Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
JOURNAL (E-mail:hattori@hgsc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
JOURNAL Fax:0427-78-9561)
COMMENT This sequence is conducted by Kitasato University JST sequencing
COMMENT Laboratory as a JST sequencing team.
COMMENT Principal Investigator:Yoshiyuki Sakaki Ph.D.
COMMENT Phone: +81-3-5449-5622, Fax: +81-3-5449-5445,
COMMENT sakaki@hgsc.ims.u-tokyo.ac.jp
COMMENT Sub-Leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
COMMENT sequence is submitted by Human Genome Sequencing in ALIS project of
COMMENT JST
COMMENT Japan Science and Technology Corporation (JST)
COMMENT 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0028 Japan
COMMENT For further information about this sequence, including its location
COMMENT and relationship to other sequences, please visit our sequence
COMMENT archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>)
COMMENT or send email to webmaster@www-alis.tokyo.jst.go.jp'.
FEATURES
source
source 1..100000
source /organism="Homo sapiens"
source /db_xref="taxon:9606"
source /chromosome="21"
source /clone="245P17-f4M4.6"
source /map="21q22.1"
source /note="D21S65.PCR1;The location is between each flanking
source site of PCR primers."
source /db_xref="GDB:191986"
BASE COUNT 26660 a 22366 c 22356 g 28618 t
ORIGIN
Query Match 56.0%; Score 22.4; DB 9; Length 100000;
Best Local Similarity 72.5%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1 ctatttccaccatcccccttggcccttggtggg 40
Db 71243 CCATTGCTTCGCATATACAGTTGGCCCTTGAATGTTGG 71204
RESULT 13
AP000124/c
LOCUS AP000124 100000 bp DNA linear PRI 25-SEP-1999
DEFINITION Homo sapiens genomic DNA of 21q22.1, GART and AML related,
DEFINITION SLC5A3-f4A4 region, segment 7/8, complete sequence.
ACCESSION AP000124
VERSION AP000124.1 GI:4730893
KEYWORDS HTG.
SOURCE Homo sapiens DNA.
ORGANISM
ORGANISM Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Homo sapiens 817,199bp genomic DNA of 21q22.1 GART and AML region
JOURNAL Published Only in DataBase (1999) In press
JOURNAL 2 (bases 1 to 100000)
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1999) to the DBJ/EMBL/GenBank databases, Mika
JOURNAL Hirakawa, Japan Science and Technology Corporation (JST), Advanced
JOURNAL Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
JOURNAL Japan (E-mail:mika@tokyo.jst.go.jp,
JOURNAL URL:<http://www-alis.tokyo.jst.go.jp/>, Tel:81-3-5214-8491,
JOURNAL Fax:81-3-5214-8470)
COMMENT This sequence is conducted by Kitasato University JST sequencing
COMMENT Laboratory as a JST sequencing team.
COMMENT Principal Investigator:Yoshiyuki Sakaki Ph.D.
COMMENT Phone: +81-3-5449-5622, Fax: +81-3-5449-5445,

```

/chromosome="21"
/map="21q22.1"
BASE COUNT      26981 a 22407 c 22383 g 28229 t
ORIGIN

Query Match      56.0%; Score 22.4; DB 9; Length 100000;
Best Local Similarity 72.5%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ctattccaccaccatcccccttgcccttggtgtggg 40
|||||  |||||  |||||  |||||  |||||  |||||
Db 61873 CCATTGCCCTTCCCATATACAGTTGGCCCTTGAATGTTGG 61834

RESULT 15
LOCUS      AL133230      111039 bp      DNA      linear      PRI 09-MAR-2001
DEFINITION Human DNA sequence from clone RP4-530115 on chromosome 20 Contains
             the 3' end of the PTPN1 gene for protein tyrosine phosphatase,
             non-receptor type 1 (EC 3.1.3.48), the gene for a novel protein
             similar to placental protein DIFF40, an RPL36 (60S ribosomal
             protein L36) pseudogene, a novel gene, two putative novel genes,
             ESTs, STSS and GSSs, complete sequence.
ACCESSION  AL133230
VERSION    AL133230.25 GI:8574103
KEYWORDS   HTG; DIFF40; protein tyrosine phosphatase; PTPN1; ribosomal
             protein; RPL36.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Wall, M.
            Direct Submission
            Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequests@sanger.ac.uk
            On Jun 20, 2000 this sequence version replaced gi:8573762.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP4-530115 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP4-530115 is at 11039 in this
            sequence. The true left end of clone RP11-24414 is at 14859 in this
            sequence. The true right end of clone RP5-894K16 is at 101 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            RP4-530115 is from the library RPC1-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
            Location/Qualifiers
                1..111039

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP4-530115"
/clone_lib="RPC1-4"
complement(7740..8185)
/note="match: GSS: Em:AQ266384"
complement(9185..9306)
/note="match: GSS: Em:AQ338716"
complement(9859..10243)
/note="match: GSS: Em:AQ080796"
complement(12783..13127)
/note="match: GSS: Em:B40765"
13108..13565
/note="match: GSS: Em:B33390"
13108..13577
/note="match: GSS: Em:B80847"
13974..13997
/note="12 copies 2 mer to 100% conserved"
14859..15356
/note="match: GSS: Em:AQ486529"
19086..19745
/note="match: GSS: Em:AQ323112"
19450..19489
/note="20 copies 2 mer to 82% conserved"
join(<20952..21042,24558..24658,27969..28067,34106..34243,
38009..38218,38757..38918,39292..39515,40854..41049,
42281..44351)
/gene="PTPN1"
/note="match: cDNAs: Em:M33962 Em:M33689 Em:M31724
match: ESTs: Em:AW612021 Em:A1611089 Em:AA262690
Em:A1700791 Em:A1802094 Em:A1498393 Em:A1040363
Em:AA256724 Em:AA844542 Em:AA596115 Em:AL121459
Em:AA074423 Em:R06661 Em:AA518943 Em:R07707 Em:AW014414
Em:R07708 Em:R06605 Em:AA005020 Em:AU041724 Em:AJ399074
Em:A1765849 Em:AQ009790 Em:R37203 Em:A1399845 Em:N1628
Em:A1611145 Em:AA258652 Em:A1206733 Em:A1950311 Em:AAT779318
Em:A1474453 Em:A1942347 Em:H68021 Em:AW243145 Em:A1184255
Em:R14704 Em:A1324731 Em:AA029531 Em:AW260499 Em:A1152511
Em:R93274 Em:AA400535 Em:A1303591 Em:AA036759 Em:A1206753
Em:AW612072 Em:A1315563 Em:AA069973 Em:A1981440
Em:A1479001 Em:AA613453 Em:A1096725 Em:A1803199
Em:AW008263 Em:A1607996 Em:AA987294 Em:A1867676
Em:AWM161410 Em:AW103475 Em:AA381416 Em:N80379 Em:A1770122
Em:AW613999 Em:T57382 Em:AA837308 Em:A1969317 Em:T52697
Em:A1830633 Em:A1147577 Em:A1038939 Em:A1693387
Em:AA193539 Em:A1417181 Em:AA070725 Em:N54866 Em:AA400489
Em:AA929060 Em:A1233029 Em:A1913603 Em:A1569873
Em:A1823502 Em:A1751202 Em:A1470804 Em:AA243361
Em:A1590278 Em:AW357407"
/product="dj530115.1 (protein tyrosine phosphatase,
non-receptor type 1 (EC 3.1.3.48))"
/evidence="not_experimental"
20952..44351
/gene="PTPN1"
join(<20952..21042,24558..24658,27969..28067,34106..34243,
38009..38218,38757..38918,39292..39515,40854..41049,
42281..44304)
/gene="PTPN1"
/note="continued from dj894K16 (AL034429)
match: proteins: Sw:P35821 Sw:O13016 Sw:P18031 Sw:P20417"
/codon_start=1
/evidence="not_experimental"
/product="dj530115.1 (protein tyrosine phosphatase,
non-receptor type 1 (EC 3.1.3.48))"
/protein_id="CAC00618.1"
/db_xref="GI:9588103"
/gene="PTPN1"
/translation="DIRHVSDFPCRVAKLPKNKNRNVSDVSPDHSRIKLHOBND
YINASLIKMEPAORSYILTQGPLPNTCGHFWMEQKRGVVMVNRWKGSLKCAQ
YWPQKEKEMIFEDTNLKITLISEDIKSYITVRQLEENLTQTREILHHYTPWD
FGVPESFASFLNLFKRVISGLSPHGPVVVHCSAGIRSGTCLADTCLLLMDKRR
DPSSVDITKKVLEMRKFRIGLIQTADQLRFSLAVIEGAKFIMGDSVQDQWKLSE

```

FEATURES
source

DLPPPPHHPPRRPRKRIILEPHNGKCRFFPNHQWVKEETQEDKDCPKIKEKGSPLN
AAPYIGESMSQDTEVRVSGVGGSLRGAQAASPAKGPSLPEKDEDHASLYWKPFVLYNM
CVATVLTAGAYLCYRFLFNSNT"
21286. .21806
/gene="PTPN1"
/notes="match: GSS: Em:AQ701555"
29085. .29126
/notes="3 copies 14 mer 97% conserved"
29101. .29289
/notes="7 copies 27 mer 88% conserved"
29209. .29250
/notes="3 copies 14 mer 97% conserved"
29237. .29275
/notes="3 copies 13 mer 92% conserved"
29304. .29345
/notes="3 copies 14 mer 90% conserved"
29319. .29399
/notes="3 copies 27 mer 91% conserved"
29372. .29413
/notes="3 copies 14 mer 92% conserved"
30667. .30759
/notes="3 copies 31 mer 78% conserved"
31563. .31618
/notes="2 copies 28 mer 94% conserved"
complement(40239. .40772)
/note="match: GSS: Em:AQ482949"
40779. .41258
/gene="PTPN1"
/notes="match: GSS: Em:AQ597355"
41484. .41669
/gene="PTPN1"
/note="match: STS: Em:G32717"
44324. .44329
/gene="PTPN1"
44345
/gene="PTPN1"
44351
/gene="PTPN1"
complement(join(45697. .46688,52105. .52135,52624. .>52785))
/gene="dJ530115.3"
/note="second exon lacks correct 3' splice site,
Em:AV708719 doesn't cover region between 46470 and 46415
in the third exon
match: ESTs: Em:AV708719 Em:AW243478 Em:AI056877
Em:AI631908 Em:AW243363 Em:AW611735 Em:AW339953
Em:AW270879 Em:AI151129 Em:H77683 Em:AI261200 Em:H63418
Em:AI206264 Em:AI279714 Em:AI242037 Em:AW589773
Em:AW445044 Em:AW593763
/product="dJ530115.3.2 (novel protein, isoform 2
(alternative 3' UTR))"
/evidence="not_experimental"
complement(join(45697. .46688,52105. .52135,52272. .52785,
54142. .54386,54958. .55093,55729. .55852,57169. .57198))
/gene="dJ530115.3"
complement(45697)
/gene="dJ530115.3"
45721. .46105
/note="match: STS: Em:G27961"
complement(45725. .45730)
/gene="dJ530115.3"
complement(join(47391. .47481,49275. .49601))
/gene="dJ530115.4"
/note="match: ESTs: Em:H63475"
/product="dJ530115.4 (putative novel transcript)"
/evidence="not_experimental"
complement(join(47391. .47481,49275. .49601))
/gene="dJ530115.4"
48339. .48382
/note="22 copies 2 mer aa 84% conserved"
51949. .52409
/note="match: GSS: Em:AQ242694"

Query Match

56.0%; Score 22.4; DB 9; Length 111039;

* . . .

Best Local Similarity 72.5%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ctatttccccccatcccccttggtggcccttggtggtggg 40
|| ||||| ||||| ||||| || ||||| || ||||| ||
Db 23280 CTTATCCCATTCATATCCCTTTGGACCTAGGGAGG 23319

Search completed: June 4, 2002, 17:48:43
Job time: 6791 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:08:55 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 cataactgaccctactctactctggtgactgggtggg 40

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 5748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 24.2 | 60.5 | 534 | 12 | TA297H02Q |
| 2 | 23.8 | 59.5 | 217 | 10 | BF987540 |
| 3 | 23.8 | 59.5 | 687 | 12 | AG112675 |
| 4 | 22.2 | 55.5 | 334 | 10 | BF813535 |
| 5 | 22.2 | 55.5 | 390 | 12 | A2050757 |
| 6 | 22.2 | 55.0 | 520 | 12 | A2856792 |
| 7 | 21.8 | 54.5 | 559 | 10 | B1538782 |
| 8 | 21.8 | 54.5 | 636 | 10 | B1659337 |
| 9 | 21.6 | 54.0 | 230 | 9 | A1629579 |
| 10 | 21.6 | 54.0 | 343 | 9 | A1586386 |
| 11 | 21.6 | 54.0 | 350 | 9 | A1920579 |
| 12 | 21.6 | 54.0 | 421 | 9 | A1438418 |
| 13 | 21.6 | 54.0 | 422 | 9 | A1600794 |
| 14 | 21.6 | 54.0 | 426 | 9 | A1065660 |
| 15 | 21.6 | 54.0 | 437 | 9 | AW917931 |
| 16 | 21.6 | 54.0 | 462 | 10 | B1388746 |
| 17 | 21.6 | 54.0 | 474 | 9 | AW091238 |

| | | | | | | |
|----|------|------|-----|----|----------|--------------------|
| 18 | 21.6 | 54.0 | 486 | 9 | AI987339 | AI987339 660004B08 |
| 19 | 21.6 | 54.0 | 493 | 9 | AI065928 | AI065928 a144h01.x |
| 20 | 21.6 | 54.0 | 501 | 9 | AW360562 | AW360562 660035F05 |
| 21 | 21.6 | 54.0 | 510 | 9 | AI665059 | AI665059 605006C07 |
| 22 | 21.6 | 54.0 | 544 | 9 | AI691732 | AI691732 605007F07 |
| 23 | 21.6 | 54.0 | 558 | 10 | B1542879 | B1542879 949074A05 |
| 24 | 21.6 | 54.0 | 565 | 12 | TA92007Q | AL462429 T. brucei |
| 25 | 21.6 | 54.0 | 567 | 9 | AI622835 | AI622835 486105A04 |
| 26 | 21.6 | 54.0 | 569 | 9 | AI739761 | AI739761 606050E03 |
| 27 | 21.6 | 54.0 | 573 | 9 | AI795446 | AI795446 605006C07 |
| 28 | 21.6 | 54.0 | 577 | 9 | AI714407 | AI714407 606013D02 |
| 29 | 21.6 | 54.0 | 578 | 9 | AW225118 | AW225118 687056F02 |
| 30 | 21.6 | 54.0 | 581 | 10 | BM333923 | BM333923 MEST179-A |
| 31 | 21.6 | 54.0 | 583 | 9 | AI734339 | AI734339 606029G11 |
| 32 | 21.6 | 54.0 | 595 | 9 | AW331631 | AW331631 687044C07 |
| 33 | 21.6 | 54.0 | 596 | 10 | BM078367 | BM078367 MEST118-G |
| 34 | 21.6 | 54.0 | 599 | 9 | AI670159 | AI670159 605019B06 |
| 35 | 21.6 | 54.0 | 617 | 9 | AI677241 | AI677241 605050C08 |
| 36 | 21.6 | 54.0 | 679 | 9 | AI676091 | AI676091 605046D05 |
| 37 | 21.4 | 53.5 | 298 | 9 | AI154465 | AI154465 ua02h01.r |
| 38 | 21.4 | 53.5 | 301 | 10 | BF896828 | BF896828 PM3-MT011 |
| 39 | 21.4 | 53.5 | 418 | 10 | W66884 | W66884 me20b09.r1 |
| 40 | 21.4 | 53.5 | 431 | 12 | A0722584 | A0722584 HS-5236-A |
| 41 | 21.4 | 53.5 | 443 | 9 | AA125134 | AA125134 mp78c05.r |
| 42 | 21.4 | 53.5 | 466 | 9 | AI180661 | AI180661 uc47c10.r |
| 43 | 21.4 | 53.5 | 488 | 10 | BM232888 | BM232888 K0329E04 |
| 44 | 21.4 | 53.5 | 491 | 10 | BE456863 | BE456863 ut30q08.y |
| 45 | 21.4 | 53.5 | 502 | 10 | BF986408 | BF986408 QV4-GN014 |

ALIGNMENTS

RESULT 1
TA297H02Q
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 297h02, reverse sequence,
genomic survey sequence.
ACCESSION
AL489616
VERSION
AL489616.1 GI:11864911
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei.
ORGANISM
Trypanosoma brucei
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE
1 (bases 1 to 534)
AUTHORS
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE
Direct Submission
JOURNAL
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhlsanger.ac.uk
COMMENT
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at <http://www.sanger.ac.uk/projects/T-brucei/>.
Location/Qualifiers
1. .534
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="297H02"

BASE COUNT 135 a 117 c 135 g 147 t

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-CI0148-271100-277-e05&t3=2000-11-27&t4=1>)
271100-277-e05&t3=2000-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 329.

FEATURES

source

Location/Qualifiers
1. .334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0148"
/dev_stage="Adult"
/note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 73 a 95 c 105 g 60 t 1 others
ORIGIN

Query Match 55.5%; Score 22.2; DB 10; Length 334;
Best Local Similarity 77.1%; Pred. No. 2.1e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 6 cttgacctactctacctggctggactgggtggg 40
||||| ||| ||| ||||||||| ||| |||||
Db 130 CTTGAACCTCTCGCCCTGGCTGGCTGCTGATTGGG 96

RESULT

AZ050757 390 bp DNA linear GSS 09-JAN-2001
LOCUS GSSTC11608 Trypanosoma cruzi random genomic library Trypanosoma
DEFINITION cruzi genomic clone G41K6, DNA sequence.
ACCESSION AZ050757
VERSION AZ050757.3 GI:10136571
KEYWORDS GSS.
SOURCE Trypanosoma cruzi.

ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 390)
REFERENCE Aguerro,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
AUTHORS A random sequencing approach for the analysis of the trypanosoma
TITLE cruzi genome: general structure, large gene and repetitive DNA
families, and gene discovery
JOURNAL Genome Res. 10 (12), 1996-2005 (2000)
MEDLINE 20568489
COMMENT On Sep 14, 2000 this sequence version replaced gi:9371531.
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24
CP(1650) San Martin, Prov. de BS AS. Argentina
Tel: 54-11-4580-7255 ext 309

Fax: 54-11-4752-9639
Email: dsanchez@lib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see <http://genome.washington.edu>). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Seq primer: r7
Class: shotgun.
Location/Qualifiers
1. .390
/organism="Trypanosoma cruzi"
/strain="CL-Brener"
/db_xref="taxon:5693"
/clone_lib="Trypanosoma cruzi random genomic library"
/cell_type="epimastigote"
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
randomly sheared using a nebulizer and the 1 to 2 Kb range
was gel purified and cloned into the dephosphorylated
HincII site of the vector."
BASE COUNT 78 a 103 c 111 g 98 t
ORIGIN

FEATURES

source

Location/Qualifiers
1. .390
/organism="Trypanosoma cruzi"
/strain="CL-Brener"
/db_xref="taxon:5693"
/clone_lib="Trypanosoma cruzi random genomic library"
/cell_type="epimastigote"
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
randomly sheared using a nebulizer and the 1 to 2 Kb range
was gel purified and cloned into the dephosphorylated
HincII site of the vector."
BASE COUNT 78 a 103 c 111 g 98 t
ORIGIN

Query Match 55.5%; Score 22.2; DB 12; Length 390;
Best Local Similarity 77.1%; Pred. No. 2.2e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 cacttgacctactctacctggctggactgggtg 38
||||| ||| ||| ||||||| ||| |||||||
Db 237 CACTTGAACTTCTATGCTGCTGGCTGAACGGGAG 271

RESULT

AZ856792 520 bp DNA linear GSS 21-FEB-2001
LOCUS 2M0161P17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0161P17 F, DNA sequence.
ACCESSION AZ856792
VERSION AZ856792.1 GI:13048135
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 520)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0161 row: P column: 17
Seq primer: CGTTGTAAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 520.

FEATURES

source

Location/Qualifiers
1. .520
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0161P17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"


```

Query Match      54.0%; Score 21.6; DB 9; Length 350;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 5 acctgacctactctacctggcgctggactgggtggg 40
    |||||  |  || |||||  |||||  |||||  |
Db 207 ACTTGACCAGATCCTGACTGGCTGGACGGGAGCG 242

RESULT 12
LOCUS      AI438418
DEFINITION 486009G01.x2 486 - leaf primordia cDNA library from Hake lab 1999
            mays cDNA, mRNA sequence.
ACCESSION  AI438418
VERSION     AI438418.1 GI:4296188
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 421)
REFERENCE   1 (bases 1 to 421)
AUTHORS    Walbot,V.
TITLE      Zea mays.
JOURNAL     Unpublished (1999)
COMMENT     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486009 row: G column: 01.
            Location/Qualifiers
                1..421
                /organism="Zea mays"
                /cultivar="B73"
                /db_xref="taxon:4577"
                /clone_lib="486 - leaf primordia cDNA library from Hake
                lab"
                /tissue_type="leaf primordia"
                /dev_stage="p7-p11 leaf"
                /lab_host="E.coli XL1-Blue MFR"
                /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
                library."

BASE COUNT  118 a 103 c 96 g 104 t
ORIGIN
1..421
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="486 - leaf primordia cDNA library from Hake
lab"
/tissue_type="leaf primordia"
/dev_stage="p7-p11 leaf"
/lab_host="E.coli XL1-Blue MFR"
/note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
library."

Query Match      54.0%; Score 21.6; DB 9; Length 421;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 5 acctgacctactctacctggcgctggactgggtggg 40
    |||||  |  || |||||  |||||  |||||  |
Db 223 ACTTGACCAGATCCTGACTGGCTGGACGGGAGCG 258

RESULT 13
LOCUS      AI600794
DEFINITION 486068G02.x1 486 - leaf primordia cDNA library from Hake lab 1999
            mays cDNA, mRNA sequence.
ACCESSION  AI600794
VERSION     AI600794.1 GI:4609955
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 426)
REFERENCE   1 (bases 1 to 426)
AUTHORS    Schutz,K., de la Bastide,M., Gnoj,L., Habermann,K., Huang,E.N.,
            Parnell,L.D., Dedhia,N., Martienssen,R. and McCombie,W.R.
TITLE      Expressed sequence tags from Zea mays
JOURNAL     Unpublished (1998)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mcombie@cshl.org
            Plate: ag91 row: c column: 06
            Seq primer: M13 forward universal -21
            High quality sequence stop: 426.
            Location/Qualifiers
                1..426
                /organism="Zea mays"
                /cultivar="B73"
                /db_xref="taxon:4577"

```

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 422)
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 486068 row: G column: 02.
Location/Qualifiers
    1..422
    /organism="Zea mays"
    /cultivar="B73"
    /db_xref="taxon:4577"
    /clone_lib="486 - leaf primordia cDNA library from Hake
    lab"
    /tissue_type="leaf primordia"
    /dev_stage="p7-p11 leaf"
    /lab_host="E.coli XL1-Blue MFR"
    /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
    library."

BASE COUNT  117 a 97 c 96 g 112 t
ORIGIN
1..422
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="486 - leaf primordia cDNA library from Hake
lab"
/tissue_type="leaf primordia"
/dev_stage="p7-p11 leaf"
/lab_host="E.coli XL1-Blue MFR"
/note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
library."

Query Match      54.0%; Score 21.6; DB 9; Length 422;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 5 acctgacctactctacctggcgctggactgggtggg 40
    |||||  |  || |||||  |||||  |||||  |
Db 222 ACTTGACCAGATCCTGACTGGCTGGACGGGAGCG 257

RESULT 14
LOCUS      AI065660
DEFINITION ag91c06.x1 maize inflorescence immature ear library Zea mays cDNA
            clone ag91c06 3', mRNA sequence.
ACCESSION  AI065660
VERSION     AI065660.1 GI:3341067
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 426)
REFERENCE   1 (bases 1 to 426)
AUTHORS    Schutz,K., de la Bastide,M., Gnoj,L., Habermann,K., Huang,E.N.,
            Parnell,L.D., Dedhia,N., Martienssen,R. and McCombie,W.R.
TITLE      Expressed sequence tags from Zea mays
JOURNAL     Unpublished (1998)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mcombie@cshl.org
            Plate: ag91 row: c column: 06
            Seq primer: M13 forward universal -21
            High quality sequence stop: 426.
            Location/Qualifiers
                1..426
                /organism="Zea mays"
                /cultivar="B73"
                /db_xref="taxon:4577"

```

Search completed: June 4, 2002, 17:09:02
Job time: 4535 sec

```

118 a      85 c      102 g      132 t
      estimated insert size approx.1 kb

BASE COUNT
      N      G      A      C      T
      0      0      0      0      0

Query Match      54.0%;      Score 21.6;      DB 9;      Length 437;
Best Local Similarity 85.7%;      Pred. No. 3.8e+02;
Matches 24;      Conservative 4;      Mismatches 0;      Indels 0;      Gaps 0;

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:47:11 : Search time 116.71 Seconds
(without alignments)
84.186 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 catacattgaccctactctacctggctggctgactgggtggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| C 1 | 20.2 | 50.5 | 359 | 4 | US-09-085-199B-30 |
| C 2 | 20.2 | 50.5 | 1164 | 4 | US-09-085-199B-1 |
| C 3 | 20.2 | 50.5 | 4796 | 4 | US-09-085-199B-3 |
| 4 | 19.4 | 48.5 | 56516 | 2 | US-08-996-306-1 |
| 5 | 19.4 | 48.5 | 56516 | 4 | US-09-338-907-1 |
| 6 | 19.4 | 48.5 | 56516 | 4 | US-09-218-207-1 |
| 7 | 19.4 | 48.5 | 56520 | 4 | US-09-338-907-179 |
| 8 | 19.4 | 48.5 | 56520 | 4 | US-09-218-207-179 |
| 9 | 19.2 | 48.0 | 2389 | 1 | US-08-123-161A-13 |
| 10 | 19.2 | 48.0 | 2389 | 1 | US-08-483-278-13 |
| 11 | 19 | 47.5 | 969 | 3 | US-09-188-930-11 |
| 12 | 19 | 47.5 | 2515 | 1 | US-08-061-465-3 |
| 13 | 19 | 47.5 | 4002 | 2 | US-08-231-193A-53 |
| 14 | 19 | 47.5 | 4002 | 2 | US-08-486-273A-53 |
| 15 | 19 | 47.5 | 4002 | 3 | US-08-480-474-53 |
| 16 | 19 | 47.5 | 4002 | 3 | US-08-940-086A-53 |
| 17 | 19 | 47.5 | 4002 | 4 | US-08-940-035A-53 |
| 18 | 19 | 47.5 | 4017 | 2 | US-08-231-193A-49 |
| 19 | 19 | 47.5 | 4017 | 2 | US-08-486-273A-49 |
| 20 | 19 | 47.5 | 4017 | 3 | US-08-480-474-49 |
| 21 | 19 | 47.5 | 4017 | 3 | US-08-940-086A-49 |
| 22 | 19 | 47.5 | 4017 | 4 | US-08-940-035A-49 |
| 23 | 19 | 47.5 | 4053 | 2 | US-08-231-193A-47 |
| 24 | 19 | 47.5 | 4053 | 2 | US-08-486-273A-47 |
| 25 | 19 | 47.5 | 4053 | 3 | US-08-480-474-47 |
| 26 | 19 | 47.5 | 4053 | 3 | US-08-940-086A-47 |
| 27 | 19 | 47.5 | 4053 | 4 | US-08-940-035A-47 |

| | | | | | | |
|----|------|------|-------|---|-------------------|-------------------|
| 28 | 19 | 47.5 | 4068 | 2 | US-08-231-193A-5 | Sequence 5, Appli |
| 29 | 19 | 47.5 | 4068 | 2 | US-08-486-273A-5 | Sequence 5, Appli |
| 30 | 19 | 47.5 | 4068 | 3 | US-08-480-474-5 | Sequence 5, Appli |
| 31 | 19 | 47.5 | 4068 | 3 | US-08-940-086A-5 | Sequence 5, Appli |
| 32 | 19 | 47.5 | 4068 | 4 | US-08-940-035A-5 | Sequence 5, Appli |
| 33 | 19 | 47.5 | 4077 | 2 | US-08-231-193A-51 | Sequence 51, Appl |
| 34 | 19 | 47.5 | 4077 | 2 | US-08-486-273A-51 | Sequence 51, Appl |
| 35 | 19 | 47.5 | 4077 | 3 | US-08-480-474-51 | Sequence 51, Appl |
| 36 | 19 | 47.5 | 4077 | 3 | US-08-940-086A-51 | Sequence 51, Appl |
| 37 | 19 | 47.5 | 4077 | 4 | US-08-940-035A-51 | Sequence 51, Appl |
| 38 | 19 | 47.5 | 4092 | 2 | US-08-231-193A-45 | Sequence 45, Appl |
| 39 | 19 | 47.5 | 4092 | 2 | US-08-486-273A-45 | Sequence 45, Appl |
| 40 | 19 | 47.5 | 4092 | 3 | US-08-480-474-45 | Sequence 45, Appl |
| 41 | 19 | 47.5 | 4092 | 4 | US-08-940-086A-45 | Sequence 45, Appl |
| 42 | 19 | 47.5 | 6828 | 1 | US-08-061-465-1 | Sequence 11, Appl |
| 43 | 19 | 47.0 | 1149 | 4 | US-09-434-774-11 | Sequence 1, Appl |
| 44 | 18.8 | 47.0 | 72928 | 3 | US-09-009-913-1 | Sequence 1, Appl |
| 45 | 18.8 | 47.0 | 72928 | 3 | US-09-009-913-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-09-085-199B-30/c
; Sequence 30, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Huq, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS DOS 5.0
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/085,199B
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Larson, Marina T.
REGISTRATION NUMBER: 32038
REFERENCE/DOCKET NUMBER: UBC.P-013US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (970) 668-2050
TELEFAX: (970) 668-2052
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 359
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: human
FEATURE:
OTHER INFORMATION: exon 15 of HIP1
US-09-085-199B-30

Query Match 50.5%; Score 20.2; DB 4; Length 359;
Best Local Similarity 75.8%; Pred. No. 14;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctgggctgactgg 35
||||| ||| ||||| ||||| |||||
Db 188 TCTCTCGTTCCTCAAAATCTACCTGGGCTTCTCTGG 156

RESULT 2

US-09-085-199B-1/c
; Sequence 1, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Hug, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE: cDNA for Huntingtin-interacting protein
; US-09-085-199B-1

Query Match 50.5%; Score 20.2; DB 4; Length 1164;
Best Local Similarity 75.8%; Pred. No. 18;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctgggctgactgg 35
||||| ||| ||||| ||||| |||||
Db 778 TCTCTCGTTCCTCAAAATCTACCTGGGCTTCTCTGG 746

RESULT 3

US-09-085-199B-3/c

; Sequence 3, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Hug, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4796
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE:
; OTHER INFORMATION: cDNA for Huntington-interacting protein
; US-09-085-199B-3

Query Match 50.5%; Score 20.2; DB 4; Length 4796;
Best Local Similarity 75.8%; Pred. No. 23;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctgggctgactgg 35
||||| ||| ||||| ||||| |||||
Db 1486 TCTCTCGTTCCTCAAAATCTACCTGGGCTTCTCTGG 1454

RESULT 4

US-08-996-306-1
; Sequence 1, Application US/08996306
; Patent No. 5945522
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Chumakov, Ilya
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Hougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway


```

; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 56516
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1629..1870
; OTHER INFORMATION: Identification method Proscan
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1998..2000
; OTHER INFORMATION: potential start codon
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2031..2033
; OTHER INFORMATION: ATG
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11694..14332
; OTHER INFORMATION: Tyr Phos
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11930..11947
; OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12057..12103
; OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12339..12358
; OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13547..13564
; OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
; FEATURE:
; NAME/KEY: allele
; LOCATION: 13657..13703
; OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13962..13981
; OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23717..23832
; OTHER INFORMATION: exon 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25571..25660
; OTHER INFORMATION: exon 4
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 34216..34234
; OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
; FEATURE:
; NAME/KEY: allele
; LOCATION: 34469..34515
; OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 34625..34645
; OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement

```

[illegible]

Query Match 48.5%; Score 19.4; DB 4; Length 56516;
Best Local Similarity 79.3%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 6; Indels 0;

[illegible]

RESULT

```

US-09-218-207-1
; Sequence 1, Application US/09218207
; Patent No. 6346331
; GENERAL INFORMATION:
;
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilyia, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GNSSET.018CP1
; CURRENT APPLICATION NUMBER: US/09/218,207
; CURRENT FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09

```

```
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 1
;   TYPE: DNA
;   LENGTH: 56516
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: 1629..1870
;   OTHER INFORMATION: identification method Proscan
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1998..2000
;   OTHER INFORMATION: potential start codon
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 2001..2216
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2031..2033
;   OTHER INFORMATION: ATG
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 11694..14332
;   OTHER INFORMATION: Tyr Phos
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 11930..11947
;   OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12057..12103
;   OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 12339..12358
;   OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 13547..13564
;   OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 13657..13703
;   OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 13962..13981
;   OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18196..18265
;   OTHER INFORMATION: exon 2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 23717..23832
;   OTHER INFORMATION: exon 3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25571..25660
;   OTHER INFORMATION: exon 4
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 34216..34234
;   OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 34469..34515
;   OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 34625..34645
```

```
; OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 34669..34759
;   OTHER INFORMATION: exon 5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 40688..40846
;   OTHER INFORMATION: exon 6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 48070..48193
;   OTHER INFORMATION: exon 7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 50182..54523
;   OTHER INFORMATION: exon 8
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51149..51168
;   OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 51448..51494
;   OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51482..51499
;   OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51596..51613
;   OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 51612..51658
;   OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51996..52015
;   OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 54445..54450
;   OTHER INFORMATION: AATAAA
;   US-09-218-207-1

Query Match      48.5%; Score: 19.4; DB 4; Length 56516;
Best Local Similarity 79.3%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 12 cctactctacctggcgtggactgggtggg 40
    ||||| ||||| ||||| |||||
Db 2735 ccaactttaactggcgttgatgtgtggg 2763

RESULT 7
US-09-338-907-179
; Sequence 179, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: PROSTATE CANCER GENE
; FILE REFERENCE: GENSET.18CP1CP
; CURRENT APPLICATION NUMBER: US/09/338,907
; CURRENT FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
```

;
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 09/218, 207
; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 179
; LENGTH: 56520
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23716..23831
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25570..25659
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34668..34758
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 40685..40843
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 48067..48190
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 50179..54519
; OTHER INFORMATION: exon8
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 54493..54498
; OTHER INFORMATION: AATAAA
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1991..2008
; OTHER INFORMATION: upstream amplification primer 5-63
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 2505..2525
; OTHER INFORMATION: downstream amplification primer 5-63 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4091..4111
; OTHER INFORMATION: downstream amplification primer 99-622
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4528..4546
; OTHER INFORMATION: upstream amplification primer 99-622 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5475..5495
; OTHER INFORMATION: downstream amplification primer 99-621
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5927..5947
; OTHER INFORMATION: upstream amplification primer 99-621 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8127..8144
; OTHER INFORMATION: downstream amplification primer 99-619

;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8560..8578
; OTHER INFORMATION: upstream amplification primer 99-619 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11622..11639
; OTHER INFORMATION: upstream amplification primer 4-76
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12013..12037
; OTHER INFORMATION: downstream amplification primer 4-76 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11930..11947
; OTHER INFORMATION: upstream amplification primer 4-77
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12339..12358
; OTHER INFORMATION: downstream amplification primer 4-77 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12915..12932
; OTHER INFORMATION: upstream amplification primer 4-71
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13317..13334
; OTHER INFORMATION: downstream amplification primer 4-71 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13216..13233
; OTHER INFORMATION: upstream amplification primer 4-72
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13617..13636
; OTHER INFORMATION: downstream amplification primer 4-72 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13547..13564
; OTHER INFORMATION: upstream amplification primer 4-73
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13962..13981
; OTHER INFORMATION: downstream amplification primer 4-73 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 15994..16011
; OTHER INFORMATION: downstream amplification primer 99-610
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 16463..16480
; OTHER INFORMATION: upstream amplification primer 99-610 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 17304..17324
; OTHER INFORMATION: downstream amplification primer 99-609
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 17814..17832
; OTHER INFORMATION: upstream amplification primer 99-609 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 18008..18027
; OTHER INFORMATION: upstream amplification primer 4-90
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 18423..18442
; OTHER INFORMATION: downstream amplification primer 4-90 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 18699..18716
; OTHER INFORMATION: downstream amplification primer 99-607
; FEATURE:

```
; NAME/KEY: primer_bind
; LOCATION: 19164..19182
; OTHER INFORMATION: upstream amplification primer 99-607 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 22589..22609
; OTHER INFORMATION: downstream amplification primer 99-602
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 23111..23129
; OTHER INFORMATION: upstream amplification primer 99-602 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 25098..25118
; OTHER INFORMATION: downstream amplification primer 99-600
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 25657..25674
; OTHER INFORMATION: upstream amplification primer 99-600 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 26537..26557
; OTHER INFORMATION: downstream amplification primer 99-598
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 27022..27040
; OTHER INFORMATION: upstream amplification primer 99-598 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 32262..32281
; OTHER INFORMATION: downstream amplification primer 99-592
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 32823..32841
; OTHER INFORMATION: upstream amplification primer 99-592 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 34215..34233
; OTHER INFORMATION: upstream amplification primer 99-217
; FEATURE:
```

Query Match 48.5%; Score 19.4; DB 4; Length 56520;

Best Local Similarity 79.3%; Pred. No. 71;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Oy 12 cctactctacctggcgactgggtggtg 40
    || ||| || ||||| || |||||
Db 2735 ccaacttaactgggcttgatgtgtg 2763
```

RESULT 8

```
US-09-218-207-179
; Sequence 179, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilyu, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GENSET.018CPI
; CURRENT APPLICATION NUMBER: US/09/218,207
; CURRENT FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 179
; LENGTH: 56520
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23716..23831
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25570..25659
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34668..34758
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 40685..40843
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 48067..48190
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 50179..54519
; OTHER INFORMATION: exon8
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 54493..54498
; OTHER INFORMATION: AATAAA
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1991..2008
; OTHER INFORMATION: upstream amplification primer 5-63
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 2505..2525
; OTHER INFORMATION: downstream amplification primer 5-63 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4091..4111
; OTHER INFORMATION: downstream amplification primer 99-622
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4528..4546
; OTHER INFORMATION: upstream amplification primer 99-622 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5475..5495
; OTHER INFORMATION: downstream amplification primer 99-621
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5927..5947
; OTHER INFORMATION: upstream amplification primer 99-621 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8127..8144
; OTHER INFORMATION: downstream amplification primer 99-619
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8560..8578
; OTHER INFORMATION: upstream amplification primer 99-619 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11622..11639
; OTHER INFORMATION: upstream amplification primer 4-76
; FEATURE:
```

```
NAME/KEY: primer_bind
LOCATION: 12018..12037
OTHER INFORMATION: downstream amplification primer 4-76 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12915..12932
OTHER INFORMATION: upstream amplification primer 4-71
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13317..13334
OTHER INFORMATION: downstream amplification primer 4-71 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13216..13233
OTHER INFORMATION: upstream amplification primer 4-72
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13617..13636
OTHER INFORMATION: downstream amplification primer 4-72 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
```

```
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32262..32281
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
OTHER INFORMATION: upstream amplification primer 99-592 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34624..34644
OTHER INFORMATION: upstream amplification primer 99-217

Query Match      48.5%   Score 19.4;   DB 4;   Length 56520;
Best Local Similarity 79.3%;   Pred. No. 71;
Matches 23;   Conservative 0;   Mismatches 6;   Indels 0;   Gaps 0;

Qy 12 cctactctacctggcgctgactgggtggg 40
    ||||| ||||| ||||| |||||
Db 2735 ccaacttaactggcgctggaatgtggg 2763

RESULT 9
US-08-123-161A-13
; Sequence 13, Application US/08123161A
; Patent No. 5449616
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Roberts, Steven L.
; APPLICANT: Anderson, Richard D.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123.161A
; FILING DATE: 16-SEP-93
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

; APPLICATION NUMBER: 07/946,234
; FILING DATE: 14-SEP-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 106..1620
US-08-123-161A-13

Query Match 48.0%; Score 19.2; DB 1; Length 2389;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctgggtgactg 34
||||| ||||| | ||||| |||||
Db 1786 TCACGTGGACCTGCCGCTGCTGGGTGGCCTG 1817

RESULT 10
US-08-483-278-13
; Sequence 13, Application US/08483278
; Patent No. 5686073
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Ervasti, James M.
; APPLICANT: Levellie, Cynthia J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,278
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,161
; FILING DATE: 16-SEP-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 106..1620
US-08-483-278-13

Query Match 48.0%; Score 19.2; DB 1; Length 2389;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctgggtgactg 34
||||| ||||| | ||||| |||||
Db 1786 TCACGTGGACCTGCCGCTGCTGGGTGGCCTG 1817

RESULT 11
US-09-188-930-11
; Sequence 11, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 969
; TYPE: DNA
; ORGANISM: mouse
US-09-188-930-11

Query Match 47.5%; Score 19; DB 3; Length 969;
Best Local Similarity 71.4%; Pred. No. 51;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 6 cttagaccctactctacctgggtgactggtggg 40
||| ||||| ||||| ||||| ||||| |||||
Db 432 ctagtcccaacttccctggcctggcctggtgtg 466

RESULT 12
US-08-061-465-3
; Sequence 3, Application US/08061465
; Patent No. 5625049
; GENERAL INFORMATION:
; APPLICANT: Monroe, Stephan S.
; APPLICANT: Glass, Roger I.
; APPLICANT: Koopmans, Marion
; APPLICANT: Jiang, Baoming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING HUMAN ASTROVIRUS
; TITLE OF INVENTION: SEROTYPE II AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 127 Peachtree Street, N.E.
; STREET: Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

LOCATION: 189...3833
US-08-486-273A-53

Query Match 47.5%; Score 19; DB 2; Length 4002;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 3 tcaattgaccctactctacctgggtgactgggt 37
||||| | | | | | | | | | | | | | | |
Db 220 TCACCTCGCTCTTCGGTGCCTGGCAGGCTGGGT 254

RESULT 15
US-08-480-474-53
; Sequence 53, Application US/08480474
; Patent No. 6033865
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA ENCODING
; TITLE OF INVENTION: SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.474
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9382B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 189...3833
US-08-480-474-53

Query Match 47.5%; Score 19; DB 3; Length 4002;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 3 tcaattgaccctactctacctgggtgactgggt 37
||||| | | | | | | | | | | | | | | |
Db 220 TCACCTCGCTCTTCGGTGCCTGGCAGGCTGGGT 254

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:56:19 ; Search time 521.5 Seconds
(without alignments)
131.690 Million-cell updates/sec

Title: US-09-721-543A-15

Perfect score: 40

Sequence: 1 ctatttccaccatcccttggcccttgggtggtggtg 40

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 40 | 100.0 | 40 | 22 AAD10601 | DNA ligand #7 for |
| 2 | 29 | 72.5 | 40 | 22 AAD10602 | DNA ligand #8 for |
| C 3 | 22 | 55.0 | 8439 | 22 AAL37594 | Human musculoskele |
| C 4 | 22 | 55.0 | 9875 | 22 AAL37595 | Human musculoskele |
| C 5 | 21.4 | 53.5 | 455 | 22 ABA42775 | Human breast cell |
| C 6 | 21.4 | 53.5 | 455 | 22 ABA47902 | Human breast cell |
| C 7 | 21.4 | 53.5 | 455 | 22 ABA53200 | Human foetal liver |
| C 8 | 21.4 | 53.5 | 455 | 22 ABA5792 | Human foetal liver |
| C 9 | 21.4 | 53.5 | 455 | 22 ABA22976 | Probe #1442 for ge |

| | | | | | |
|------|------|------|--------|-------------|---------------------|
| C 10 | 21.4 | 53.5 | 455 | 22 ABA33881 | Probe #11347 for g |
| C 11 | 21.4 | 53.5 | 455 | 22 AAK01450 | Human brain expres |
| C 12 | 21.4 | 53.5 | 455 | 22 AAK14197 | Human brain expres |
| C 13 | 21.4 | 53.5 | 455 | 22 AAK26895 | Human bone marrow |
| C 14 | 21.4 | 53.5 | 455 | 22 AAK39928 | Human bone marrow |
| C 15 | 21.4 | 53.5 | 455 | 22 AAI11524 | Probe #1457 for ge |
| C 16 | 21.4 | 53.5 | 455 | 22 AAI20737 | Probe #10670 for g |
| C 17 | 21.4 | 53.5 | 455 | 22 AAI33801 | Probe #1487 used t |
| C 18 | 21.4 | 53.5 | 455 | 22 AAI45954 | Probe #14640 used |
| C 19 | 21.4 | 53.5 | 455 | 22 AAI01440 | Probe #1431 used t |
| C 20 | 21.4 | 53.5 | 455 | 22 AAI06437 | Probe #6428 used t |
| C 21 | 21.4 | 53.5 | 549 | 21 AAC93601 | Cat flea head and |
| C 22 | 21.4 | 53.5 | 12600 | 24 AAD25770 | Human APOBEC2 gene |
| C 23 | 21.4 | 53.5 | 12600 | 24 AAD25817 | Human APOBEC2 gene |
| C 24 | 21.4 | 53.5 | 13294 | 22 AAK85255 | Human immune/haema |
| C 25 | 21 | 52.5 | 2004 | 22 AAI61050 | Human polynucleoti |
| C 26 | 21 | 52.5 | 19038 | 22 AAF30492 | Human PAK5 gene. |
| C 27 | 20.8 | 52.0 | 534 | 21 AAF10046 | Fusarium venenatum |
| C 28 | 20.8 | 52.0 | 1236 | 23 AAS53753 | Helicobacter pylori |
| C 29 | 20.8 | 52.0 | 1350 | 19 AAX14250 | H. pylori GHPO 144 |
| C 30 | 20.8 | 52.0 | 5903 | 23 ABL23386 | Drosophila melanog |
| C 31 | 20.8 | 52.0 | 12710 | 22 AAK81012 | Human immune/haema |
| C 32 | 20.8 | 52.0 | 25772 | 22 AAK81332 | Human immune/haema |
| C 33 | 20.6 | 51.5 | 550 | 22 AAK36788 | Human colon cancer |
| C 34 | 20.6 | 51.5 | 699 | 22 AAL15715 | Human breast cancer |
| C 35 | 20.6 | 51.5 | 1087 | 21 AAF21828 | Human breast and o |
| C 36 | 20.6 | 51.5 | 1581 | 17 AAT31798 | Erythrobacter long |
| C 37 | 20.4 | 51.0 | 362 | 22 AAK74501 | Human immune/haema |
| C 38 | 20.4 | 51.0 | 373 | 22 AAK59097 | Human immune/haema |
| C 39 | 20.4 | 51.0 | 576 | 21 AAA08747 | Murine interferon |
| C 40 | 20.4 | 51.0 | 16747 | 22 AAL36989 | Human musculoskele |
| C 41 | 20.4 | 51.0 | 16747 | 22 AAK80326 | Human immune/haema |
| C 42 | 20.4 | 51.0 | 16747 | 22 AAK86604 | Human immune/haema |
| C 43 | 20.4 | 51.0 | 16747 | 22 AAK90152 | Human digestive sy |
| C 44 | 20.4 | 51.0 | 56390 | 23 ABL22618 | Drosophila melanog |
| C 45 | 20.4 | 51.0 | 240825 | 22 AAF24497 | Human PG-3 gene. |

ALIGNMENTS

RESULT 1
AAD10601
ID AAD10601 standard; DNA; 40 BP.
XX
AC AAD10601;
XX
DT 24-SEP-2001 (first entry)
XX
DE DNA ligand #7 for human cytomegalovirus.
XX
KW Ligand: human cytomegalovirus; HCMV infection; herpetic ulcer;
KW pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW herpesvirus infection; lesion; ds.
XX
OS Human cytomegalovirus.
OS Synthetic.
XX
PN WO200138341-A1.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32259.
XX
PR 24-NOV-1999; 99US-0167509.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Liu F, Wang J, Jiang H;
XX
DR WPI: 2001-367658/38.
XX
PT Polynucleotide ligands useful as anti-viral agents for the treatment of

PT herpesvirus infections e.g. primary or chronic infections with human
 XX cytomegalovirus -

PS Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
 CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
 CC or block HCMV infection in target cells. These ligands are used as
 CC anti-viral agents for the treatment of herpesvirus infections (e.g.
 CC lesions and herpetic ulcers) and particularly HCMV associated
 CC primary or chronic infections such as pneumonia, gastroenteritis
 CC and chorioretinitis. The present sequence is DNA ligand which bind
 CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 4 A; 14 C; 10 G; 12 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;

Best Local Similarity 100.0%; Pred. No. 1.3e-06;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatttccaccatattcccttggcccttgggtgtggg 40

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 ctatttccaccatattcccttggcccttgggtgtggg 40

RESULT 2

AAD10602

ID AAD10602 standard; DNA; 40 BP.

XX AC AAD10602;

XX DT 24-SEP-2001 (first entry)

XX DNA ligand #8 for human cytomegalovirus.

XX Ligand; human cytomegalovirus; HCMV infection; herpetic ulcer;
 KW pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
 KW herpesvirus infection; lesion; ds.

XX OS Human cytomegalovirus.

XX OS Synthetic.

PH Key Location/Qualifiers

FT misc_feature 13
 FT /*tag- a
 FT /note- "This residue is absent in the sequence
 FT shown as SEQ ID NO:16 in the sequence listing
 FT of the specification"

FT misc_feature 24..25
 FT /*tag- b
 FT /note- "This residue is absent in the sequence
 FT shown as SEQ ID NO:16 in the sequence listing
 FT of the specification"

XX WO200138341-A1.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000WO-US32259.

XX 24-NOV-1999; 99US-0167509.

XX (REGC) UNIV CALIFORNIA.

XX Liu F, Wang J, Jiang H;

XX WPI; 2001-367658/38.

XX Polynucleotide ligands useful as anti-viral agents for the treatment of
 PT herpesvirus infections e.g. primary or chronic infections with human
 PT cytomegalovirus -

PS Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
 CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
 CC or block HCMV infection in target cells. These ligands are used as
 CC anti-viral agents for the treatment of herpesvirus infections (e.g.
 CC lesions and herpetic ulcers) and particularly HCMV associated
 CC primary or chronic infections such as pneumonia, gastroenteritis
 CC and chorioretinitis. The present sequence is DNA ligand which bind
 CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 5 A; 15 C; 6 G; 11 T; 3 other;

Query Match 72.5%; Score 29; DB 22; Length 40;

Best Local Similarity 80.0%; Pred. No. 0.027;

Matches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ctatttccaccatattcccttggcccttgggtgtggg 40

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 ctatttccaccatattcccttggcccttgggtgtggg 40

RESULT 3

AAL37594/c

ID AAL37594 standard; DNA; 8439 BP.

XX AC AAL37594;

XX DT 08-JAN-2002 (first entry)

XX Human musculoskeletal system related polynucleotide SEQ ID NO 3959.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein;
 KW musculoskeletal system; ds.

XX OS Homo sapiens.

XX WO20015367-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214896.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-451937/48.
XX
DR
XX
XX
PT Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
XX
XX Example 2; SEQ ID NO 3959; 781pp + Sequence Listing; English.
PS
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 8439 BP; 1748 A; 2333 C; 2307 G; 2051 T; 0 other;
SQ

Query Match 55.0%; Score 22; DB 22; Length 8439;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 cccaccatcccttggcccttgggtg 36
||||| ||||| ||||| ||
Db 5444 CCACCACATCCCTGGGCGCTGAGATG 5415

RESULT 4

AAL37595/c

ID AAL37595 standard; DNA; 9875 BP.

XX

XX

AC

AA

DT

XX

DE

XX

KW

KW

KW

KW

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

17

31

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR


```
ID  ABA47902 standard; DNA: 455 BP.
XX  ABA47902;
AC
XX
XX  01-FEB-2002 (first entry)
DT
XX
XX  Human breast cell single exon nucleic acid probe #6597.
DE
XX  Human; microarray; single exon probe; gene expression; breast;
KW  disease; cancer; ss.
XX
XX  Homo sapiens.
OS
XX
XX  WO200157271-A2.
PN
XX
XX  09-AUG-2001.
PD
XX
XX  30-JAN-2001; 2001WO-US00662.
PF
XX
XX  04-FEB-2000; 2000US-0180312.
PR
XX  26-MAY-2000; 2000US-0207456.
PR
XX  30-JUN-2000; 2000US-0608408.
PR
XX  03-AUG-2000; 2000US-0632366.
PR
XX  21-SEP-2000; 2000US-0234687.
PR
XX  27-SEP-2000; 2000US-0236359.
PR
XX  04-OCT-2000; 2000US-0024263.
PR
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX  WPI; 2001-496933/54.
DR
XX
XX  New spatially-addressable set of single exon nucleic acid probes,
PT  useful for measuring gene expression in sample derived from human
PT  breast, comprises number of single exon nucleic acid probes -
PT
XX
XX  Claim 4; SEQ ID NO 6597; 327pp + sequence listing; English.
PS
XX
XX  The invention relates to a spatially-addressable set of single exon
CC  nucleic acid probes for measuring gene expression in a sample derived
CC  from human breast and BT 474 cells. The method involves contacting
CC  the probes with a collection of detectably labelled nucleic acids
CC  derived from mRNA of human breast, and then measuring the label
CC  bound to each probe of the microarray. The probes are useful for
CC  verifying the expression of regions of genomic DNA predicted to
CC  encode proteins. They are useful for gene discovery, and for
CC  determining predisposition and/or prognosing breast disease. Gene
CC  expression analysis is useful for assessing the toxicity of chemical
CC  agents on cells. The microarray of this invention presents a far greater
CC  diversity of probes for measuring gene expression, with far less bias
CC  than expressed sequence tag microarrays. The method is suitable for
CC  rapid production of functional information from genomic sequence. The
CC  present sequence is a single exon nucleic acid probe of the invention.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
CC  from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
SQ

Query Match      53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY  2 tatttcacccatcccttggtggcccttggtgtggg 40
    | ||||| ||||| || ||||| || ||||| ||
Db   90 TGTTCACCCAAATCCCTAGCTCCACTGGGGAGAG 52

RESULT 7
ABA53200/c
ID  ABA53200 standard; DNA: 455 BP.
XX
XX  ABA53200;
AC
XX
XX  01-FEB-2002 (first entry)
DT
XX
XX  Human foetal liver single exon nucleic acid probe #1505.
DE
XX  Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW
XX
XX  Homo sapiens.
OS
XX
XX  WO200157277-A2.
PN
XX
XX  09-AUG-2001.
PD
XX
XX  30-JAN-2001; 2001WO-US00669.
PF
XX
XX  04-FEB-2000; 2000US-0180312.
PR
XX  26-MAY-2000; 2000US-0207456.
PR
XX  30-JUN-2000; 2000US-0608408.
PR
XX  03-AUG-2000; 2000US-0632366.
PR
XX  21-SEP-2000; 2000US-0234687.
PR
XX  27-SEP-2000; 2000US-0236359.
PR
XX  04-OCT-2000; 2000US-0024263.
PR
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX  WPI; 2001-483447/52.
DR
XX
XX  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human foetal liver -
PT
XX
XX  Claim 1; SEQ ID NO 1505; 539pp + sequence listing; English.
PS
XX
XX  The invention relates to a single exon nucleic acid probe for
CC  measuring human gene expression in a sample derived from human foetal
CC  liver. The single exon nucleic acid probes may be used for predicting,
CC  measuring and displaying gene expression in samples derived from human
CC  foetal liver. The present sequence is a single exon nucleic acid
CC  probe of the invention.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
CC  from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
SQ

Query Match      53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY  2 tatttcacccatcccttggtggcccttggtgtggg 40
    | ||||| ||||| || ||||| || ||||| ||
Db   90 TGTTCACCCAAATCCCTAGCTCCACTGGGGAGAG 52

RESULT 8
ABA65792/c
ID  ABA65792 standard; DNA: 455 BP.
XX
XX  ABA65792;
AC
XX
XX  01-FEB-2002 (first entry)
DT
XX
XX  Human foetal liver single exon nucleic acid probe #14097.
DE
XX  Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW
XX
XX  Homo sapiens.
OS
XX
XX  WO200157277-A2.
PN
```

```
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 14097; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 tatttccaccatcccttggcccttgggtgtgg 40
Db | ||||| ||||| || ||||| || ||||| ||
90 TGTTCCTCCACCCAAATTCCTAGTCCACTGGGGGAGAG 52

RESULT 9
ABA22976/c
ID ABA22976 standard; DNA; 455 BP.
XX AC ABA22976;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #1442 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 14097; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 tatttccaccatcccttggcccttgggtgtgg 40
Db | ||||| ||||| || ||||| || ||||| ||
90 TGTTCCTCCACCCAAATTCCTAGTCCACTGGGGGAGAG 52

RESULT 10
ABA32881/c
ID ABA32881 standard; DNA; 455 BP.
XX AC ABA32881;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #11347 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID NO 1442; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
```

```
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID NO 1442; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 tatttccaccatcccttggcccttgggtgtgg 40
Db | ||||| ||||| || ||||| || ||||| ||
90 TGTTCCTCCACCCAAATTCCTAGTCCACTGGGGGAGAG 52

RESULT 10
ABA32881/c
ID ABA32881 standard; DNA; 455 BP.
XX AC ABA32881;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #11347 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID NO 1442; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
```

DR WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 4; SEQ ID NO 11347; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 tatttcacccatcccttggtggcccttggtgtggg 40
I|||||I|||||I|||||I|||||I|||||I|||||I
DB 90 TGTTCACCCAAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 11
AAK01450/c
ID AAK01450 standard; DNA: 455 BP.
XX
AC AAK01450;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 1441.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 1441; 650pp + Sequence Listing; English.
XX
XX Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
XX
XX Query Match 53.5%; Score 21.4; DB 22; Length 455;

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 tatttcacccatcccttggtggcccttggtgtggg 40
I|||||I|||||I|||||I|||||I|||||I|||||I
DB 90 TGTTCACCCAAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 12
AAK14197/c
ID AAK14197 standard; DNA: 455 BP.
XX
AC AAK14197;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 14188.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 14188; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;

| | | |
|--|---|---------------|
| DT | 06-NOV-2001 | (first entry) |
| XX | Human bone marrow expressed single exon probe SEQ ID NO: 14485. | |
| DE | Human bone marrow expressed exon; gene expression analysis; probe; | |
| XX | microarray; cancer; leukaemia; lymphoma; myeloma; ss. | |
| KW | Human; bone marrow expressed exon; gene expression analysis; probe; | |
| KW | microarray; cancer; leukaemia; lymphoma; myeloma; ss. | |
| XX | Homo sapiens. | |
| OS | WO200157276-A2. | |
| PN | XX | |
| XX | 09-AUG-2001. | |
| XX | XX | |
| XX | 30-JAN-2001; 2001WO-US00668. | |
| XX | XX | |
| PR | 04-FEB-2000; 2000US-0180312. | |
| PR | 26-MAY-2000; 2000US-0207456. | |
| PR | 30-JUN-2000; 2000US-0608408. | |
| PR | 03-AUG-2000; 2000US-0632366. | |
| PR | 21-SEP-2000; 2000US-0234687. | |
| PR | 27-SEP-2000; 2000US-0236359. | |
| PR | 04-OCT-2000; 2000GB-0024263. | |
| XX | XX | |
| PA | (MOLE-) MOLECULAR DYNAMICS INC. | |
| XX | XX | |
| PI | Penn SG, Hanzel DK, Chen W, Rank DR; | |
| XX | XX | |
| DR | WPI; 2001-488900/53. | |
| XX | XX | |
| PT | Human genome-derived single exon nucleic acid probes useful for | |
| PT | analyzing gene expression in human bone marrow - | |
| XX | XX | |
| PS | Example 4; SEQ ID NO: 14485; 658pp + Sequence Listing; English. | |
| XX | XX | |
| CC | The present invention provides a number of single exon nucleic acid | |
| CC | probes which are derived from genomic sequences expressed in the human | |
| CC | bone marrow. They can be used to measure gene expression in bone marrow | |
| CC | samples, which may enable the improved diagnosis and treatment of cancers | |
| CC | such as lymphoma, leukaemia and myeloma. The present sequence is one of | |
| CC | the probes of the invention. | |
| XX | XX | |
| SQ | Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other; | |
| Query Match 53.5%; Score 21.4; DB 22; Length 455; | | |
| Best Local Similarity 71.8%; Pred. No. 37; | | |
| Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0; | | |
| Qy | 2 tatttcacccatcccccttggcccttgggtgtgtggg 40 | |
| | | |
| Db | 90 TGTTTCCACCAAAATTCCTAGCTCCACAGGGGGAGAG 52 | |
| RESULT 15 | | |
| AA11524/c | | |
| ID | AA11524 standard; DNA; 455 BP. | |
| XX | XX | |
| AC | AA11524; | |
| XX | XX | |
| DT | 12-OCT-2001 (first entry) | |
| XX | XX | |
| DE | Probe #1457 for gene expression analysis in human cervical cell sample. | |
| XX | Probe; human; microarray; gene expression; cervical epithelial cell; | |
| KW | cervical cancer; ss. | |
| KW | Homo sapiens. | |
| XX | OS | |
| XX | WO200157278-A2. | |
| PN | XX | |
| XX | 09-AUG-2001. | |
| XX | XX | |
| PF | 30-JAN-2001; 2001WO-US00670. | |

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 1457; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 tatttccaccacatattcccttggcccttgggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | |
Db 90 TGTTCACCCCAAAATTCCTAGCTCCACTGGGGGAG 52

Search completed: June 4, 2002, 17:56:22
Job time: 7110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:56:14 ; Search time 521.5 Seconds
(without alignments)
131.690 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 catcaattgaccctactctactctggtggtggtggg 40

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 40 | 100.0 | 40 | 22 | AA10600 |
| 2 | 22.6 | 56.5 | 474 | 22 | AA11954 |
| 3 | 22.6 | 56.5 | 650 | 22 | AA12084 |
| 4 | 22.2 | 55.5 | 2544 | 22 | AA11782 |
| 5 | 21.2 | 53.0 | 119 | 22 | AAH3389 |
| 6 | 20.8 | 52.0 | 2687 | 22 | AA17508 |
| 7 | 20.8 | 52.0 | 3563 | 22 | AA17517 |
| 8 | 20.8 | 52.0 | 24813 | 22 | AAK70879 |
| 9 | 20.6 | 51.5 | 2113 | 22 | AA126576 |

| | | | | | | |
|----|------|------|-------|----|----------|--------------------|
| 10 | 20.4 | 51.0 | 615 | 21 | AAF04663 | Fusarium venenatum |
| 11 | 20.4 | 51.0 | 23474 | 23 | AA558511 | Propionibacterium |
| 12 | 20.4 | 51.0 | 24908 | 22 | AAK81665 | Human immune/haema |
| 13 | 20.2 | 50.5 | 359 | 21 | AAZ58767 | Human huntingtin-i |
| 14 | 20.2 | 50.5 | 601 | 24 | AA561792 | Lung small cell ca |
| 15 | 20.2 | 50.5 | 1164 | 18 | AA767186 | Huntingtin interac |
| 16 | 20.2 | 50.5 | 1164 | 21 | AAZ58745 | Human huntingtin-i |
| 17 | 20.2 | 50.5 | 4796 | 18 | AA767187 | Huntingtin interac |
| 18 | 20.2 | 50.5 | 4796 | 21 | AAZ58746 | Human huntingtin-i |
| 19 | 20.0 | 50.0 | 466 | 21 | AAZ26742 | Human secreted pro |
| 20 | 20.0 | 50.0 | 531 | 22 | AAK64603 | Human immune/haema |
| 21 | 20.0 | 50.0 | 3210 | 23 | ABL07885 | Drosophila melanog |
| 22 | 20.0 | 50.0 | 4377 | 23 | ABL07884 | Drosophila melanog |
| 23 | 19.8 | 49.5 | 509 | 23 | AA567978 | DNA encoding novel |
| 24 | 19.8 | 49.5 | 1280 | 23 | AA581491 | DNA encoding novel |
| 25 | 19.8 | 49.5 | 1371 | 23 | ABL14959 | Drosophila melanog |
| 26 | 19.8 | 49.5 | 1713 | 22 | ABL36883 | Human musculoskele |
| 27 | 19.8 | 49.5 | 1830 | 22 | AAH14834 | Human cDNA sequenc |
| 28 | 19.8 | 49.5 | 1849 | 22 | AAH64889 | Human secreted pro |
| 29 | 19.8 | 49.5 | 2086 | 23 | AA584430 | DNA encoding novel |
| 30 | 19.8 | 49.5 | 2497 | 22 | AA158171 | Human polynucleoti |
| 31 | 19.8 | 49.5 | 2511 | 22 | ABA09097 | Human HSPC264 homo |
| 32 | 19.8 | 49.5 | 3969 | 18 | AA762655 | DNA encoding prote |
| 33 | 19.8 | 49.5 | 4319 | 22 | AA105694 | Human reproductive |
| 34 | 19.8 | 49.5 | 4813 | 23 | ABL14958 | Drosophila melanog |
| 35 | 19.8 | 49.5 | 5351 | 22 | AA531466 | Human DNA for a no |
| 36 | 19.8 | 49.5 | 14176 | 22 | AA526670 | Human genomic DNA |
| 37 | 19.8 | 49.5 | 32249 | 22 | ABA17155 | Human nervous syst |
| 38 | 19.6 | 49.0 | 718 | 11 | AAQ03834 | Ovine growth hormo |
| 39 | 19.6 | 49.0 | 817 | 10 | AA903501 | Goat growth hormo |
| 40 | 19.6 | 49.0 | 1717 | 21 | AA78069 | Human cancer assoc |
| 41 | 19.6 | 49.0 | 2064 | 21 | AA651424 | Porcine BAC-P1GF2- |
| 42 | 19.6 | 49.0 | 2230 | 9 | AA81778 | Sequence of porcin |
| 43 | 19.6 | 49.0 | 2638 | 22 | ABA21082 | Human nervous syst |
| 44 | 19.6 | 49.0 | 3211 | 22 | AAH34203 | Human colon cancer |
| 45 | 19.6 | 49.0 | 3820 | 20 | AAZ25349 | Rat DRM nucleotide |

ALIGNMENTS

RESULT 1
AA10600
ID AA10600 standard; DNA; 40 BP.
XX
AC AA10600;
XX
DT 24-SEP-2001 (first entry)
XX
DE DNA ligand #6 for human cytomegalovirus.
XX
KW Ligand: human cytomegalovirus; HCMV infection; herpetic ulcer;
KW pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW herpesvirus infection; lesion; ds.
XX
OS Human cytomegalovirus.
OS Synthetic.
XX
PN WO200138341-A1.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32259.
XX
PR 24-NOV-1999; 99US-0167509.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Liu F, Wang J, Jiang H;
XX
DR WPI; 2001-367658/38.
XX
PT Polynucleotide ligands useful as anti-viral agents for the treatment of

PT herpesvirus infections e.g. primary or chronic infections with human
XX cytomegalovirus

PS Claim 8; Page 6; 60pp; English.

CC The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 6 A; 12 C; 12 G; 10 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catcactgaccctactctactctgctggctggactgggtggg 40
|||||
Db 1 catcactgaccctactctactctgctggctggactgggtggg 40
|||||

RESULT 2

AAL11954
ID AAL11954 standard; cDNA; 474 BP.

XX AAL11954;

XX 07-DEC-2001 (first entry)

XX Human breast cancer expressed polynucleotide 4411.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

XX 14-MAR-2000; 2000US-0189167.

XX 24-MAR-2000; 2000US-0192099.

XX 29-MAR-2000; 2000US-0193480.

XX 15-MAY-2000; 2000US-0205230.

XX 09-JUN-2000; 2000US-0211315.

XX 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer

XX Claim 1; Page 797-798; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.

XX Sequence 474 BP; 133 A; 118 C; 117 G; 104 T; 2 other;

Query Match 56.5%; Score 22.6; DB 22; Length 474;
Best Local Similarity 75.7%; Pred. No. 12;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 tcacttgaccctactctactctgctggctggactgggtgg 39
|||||
Db 152 tctctaaactagtctacttagctggcggtgg 188
|||||

RESULT 3

AAL20844
ID AAL20844 standard; cDNA; 650 BP.

XX AAL20844;

XX 07-DEC-2001 (first entry)

XX Human breast cancer expressed polynucleotide 13301.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

XX 14-MAR-2000; 2000US-0189167.

XX 24-MAR-2000; 2000US-0192099.

XX 29-MAR-2000; 2000US-0193480.

XX 15-MAY-2000; 2000US-0205230.

XX 09-JUN-2000; 2000US-0211315.

XX 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer
XX Claim 1; Page 2357; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.

XX Sequence 650 BP; 181 A; 155 C; 167 G; 142 T; 5 other;

Query Match 56.5%; Score 22.6; DB 22; Length 650;
Best Local Similarity 75.7%; Pred. No. 13;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 tcacttgaccctactctactctgctggctggactgggtgg 39
|||||
Db 135 tctctaaactagtctacttagctggcggtgg 171
|||||

```
RESULT 4
AAH17822
XX AAH17822 standard; cDNA; 2544 BP.
AC AAH17822;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17492.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000BP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 17492; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination
XX of the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 2544 BP; 467 A; 775 C; 740 G; 562 T; 0 other;
XX
XX Query Match 55.5%; Score 22.2; DB 22; Length 2544;
XX Best Local Similarity 77.1%; Pred. No. 22;
XX Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
XX QY 6 cttgacctactctactggcgtgactgggtggg 40
XX ||||| ||| ||| ||||| ||| |||
XX Db 1138 cttgacctctcgcctggcgtgactgattggg 1172

RESULT 5
AAH33899
XX AAH33899 standard; cDNA; 119 BP.
AC AAH33899;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:955.
DE
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
KW
XX Homo sapiens.
OS
XX WO200122920-A2.
PN
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI; 2001-235357/24.
XX
XX P-PSDB; AAG74468.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT
XX Claim 1; Page 2853; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB77789 represent sequences used in the exemplification of the
XX present invention.
XX
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 119 BP; 10 A; 47 C; 38 G; 23 T; 1 other;
XX
XX Query Match 53.0%; Score 21.2; DB 22; Length 119;
XX Best Local Similarity 72.2%; Pred. No. 36;
XX Matches 26; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
XX
XX QY 5 acctgacctactctactggcgtgactgggtggg 40
XX || ||| ||| ||| :||||||| | ||||
XX Db 39 acctgaacctctcgcctggcgtgactgggtggg 74

RESULT 6
AAH17508/c
XX AAH17508 standard; DNA; 2687 BP.
```



```
FT      /*tag= k
FT      /number= 5
FT      intron 2297..2422
FT      /*tag= l
FT      /number= 6
FT      exon 2423..3381
FT      /*tag= m
FT      /number= 6
FT      intron 3382..3563
FT      /*tag= n
FT      /number= 7
FT      /cons_splice= (5'site:NO, 3'site:NO)
XX      WO200166563-A2.
XX
XX      13-SEP-2001.
XX
XX      07-MAR-2001; 2001WO-US07265.
XX
XX      07-MAR-2000; 2000US-0187546.
XX      07-APR-2000; 2000US-0195536.
XX      06-JUN-2000; 2000US-0209840.
XX      23-JUN-2000; 2000US-0214213.
XX      17-AUG-2000; 2000US-0226448.
XX      03-JAN-2001; 2001US-0259227.
XX
XX      (SENO-) SENOMYX INC.
XX
XX      Adler JE, Zozulya S, O'Connell SM, Li X, Staszewski L;
XX      WPI: 2001-582267/65.
XX      P-PSDB; AAE10366.
XX
XX      New mammalian taste-cell-specific G protein-coupled receptor
XX      polypeptides for identifying compounds that modulate taste signaling
XX      are useful in food, to modulate the sweet taste of foods or drugs -
XX
XX      Claim 1; Page 71-72; 119pp; English.
XX
XX      The invention relates to mammalian taste-cell-specific G protein-coupled
XX      receptors, T1R and their corresponding cDNA molecules. Taste receptors,
XX      T1R are useful for screening compounds which are used to activate or
XX      modulate chemosensory transduction, such as taste sensation. The
XX      identification and isolation of novel taste receptors and taste
XX      signalling molecules allow for new methods of chemical and genetic
XX      modulation of taste transduction pathways. The taste modulating
XX      compounds are useful in pharmaceuticals and food industries to improve
XX      the taste of a variety of consumer products, or to block undesirable
XX      tastes, e.g., in certain pharmaceuticals. T1R.s are also useful in
XX      biochemical assay for identifying tastant (T1R) ligands having binding
XX      specificity for T1R involved in taste signalling. The present sequence is
XX      human taste-cell-specific G protein-coupled receptor, hT1R3 full-length
XX      genomic DNA. Human T1R3 gene is localised on chromosome 1p36.2-1p36.33.
XX
XX      Sequence 3563 BP; 545 A; 1214 C; 1166 G; 638 T; 0 other;
XX
XX      Query Match 52.0%; Score 20.8; DB 22; Length 3563;
XX      Best Local Similarity 70.0%; Pred. No. 83;
XX      Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
XX
Oy      1 catcacttgaccctactctactcctggctggactgggtggg 40
Db      |||| |||| | |||| |||| | |||| ||||
Db      925 CATCTTCTGCACACAGCGGTCCCGAGCGGCTGGTGGG 886

RESULT      8
AAK70879
ID      AAK70879 standard; DNA; 24813 BP.
XX
XX      AAK70879;
XX
XX      06-NOV-2001 (first entry)
XX
```

```
XX      Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25691.
XX      Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX      cytostatic; gene therapy; vaccine; metastasis; ds.
XX      Homo sapiens.
XX      WO200157182-A2.
XX      09-AUG-2001.
XX
XX      17-JAN-2001; 2001WO-US01354.
XX
XX      31-JAN-2000; 2000US-0179065.
XX      04-FEB-2000; 2000US-0180628.
XX      24-FEB-2000; 2000US-0184664.
XX      02-MAR-2000; 2000US-0186350.
XX      16-MAR-2000; 2000US-0189874.
XX      17-MAR-2000; 2000US-0190076.
XX      18-APR-2000; 2000US-0198123.
XX      19-MAY-2000; 2000US-0205515.
XX      07-JUN-2000; 2000US-0209467.
XX      28-JUN-2000; 2000US-0214886.
XX      30-JUN-2000; 2000US-0215135.
XX      07-JUL-2000; 2000US-0216647.
XX      07-JUL-2000; 2000US-0216880.
XX      11-JUL-2000; 2000US-0217487.
XX      11-JUL-2000; 2000US-0217496.
XX      14-JUL-2000; 2000US-0218290.
XX      26-JUL-2000; 2000US-0220963.
XX      26-JUL-2000; 2000US-0220964.
XX      14-AUG-2000; 2000US-0224518.
XX      14-AUG-2000; 2000US-0224519.
XX      14-AUG-2000; 2000US-0225213.
XX      14-AUG-2000; 2000US-0225214.
XX      14-AUG-2000; 2000US-0225266.
XX      14-AUG-2000; 2000US-0225267.
XX      14-AUG-2000; 2000US-0225268.
XX      14-AUG-2000; 2000US-0225270.
XX      14-AUG-2000; 2000US-0225447.
XX      14-AUG-2000; 2000US-0225757.
XX      14-AUG-2000; 2000US-0225758.
XX      14-AUG-2000; 2000US-0225759.
XX      18-AUG-2000; 2000US-0226279.
XX      22-AUG-2000; 2000US-0226681.
XX      22-AUG-2000; 2000US-0226688.
XX      22-AUG-2000; 2000US-0227182.
XX      23-AUG-2000; 2000US-0227009.
XX      30-AUG-2000; 2000US-0228924.
XX      01-SEP-2000; 2000US-0229287.
XX      01-SEP-2000; 2000US-0229343.
XX      01-SEP-2000; 2000US-0229344.
XX      01-SEP-2000; 2000US-0229345.
XX      05-SEP-2000; 2000US-0229509.
XX      05-SEP-2000; 2000US-0229513.
XX      06-SEP-2000; 2000US-0230437.
XX      06-SEP-2000; 2000US-0230438.
XX      08-SEP-2000; 2000US-0231242.
XX      08-SEP-2000; 2000US-0231243.
XX      08-SEP-2000; 2000US-0231244.
XX      08-SEP-2000; 2000US-0231413.
XX      08-SEP-2000; 2000US-0231414.
XX      08-SEP-2000; 2000US-0232080.
XX      08-SEP-2000; 2000US-0232081.
XX      12-SEP-2000; 2000US-0231968.
XX      14-SEP-2000; 2000US-0232397.
XX      14-SEP-2000; 2000US-0232398.
XX      14-SEP-2000; 2000US-0232399.
XX      14-SEP-2000; 2000US-0232400.
XX      14-SEP-2000; 2000US-0232401.
XX      14-SEP-2000; 2000US-0233063.
XX      14-SEP-2000; 2000US-0233064.
XX      14-SEP-2000; 2000US-0233065.
```

```
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 25691; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 24813 BP; 5290 A; 6858 C; 7656 G; 5009 T; 0 other;

Query Match 52.0%; Score 20.8; DB 22; Length 24813;
Best Local Similarity 78.1%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 8 tgaccctactctacgtgggtgactgggtgg 39
    ||||| ||||| ||| ||||| |||||
Db 8205 tgaccctgtctctctgtggcctggcgtgtgg 8236

RESULT 9
AAL26576/c
ID AAL26576 standard; cDNA; 2113 BP.
XX
XX AAL26576;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 19033.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
XX 14-MAR-2000; 2000US-0189167.
XX 24-MAR-2000; 2000US-0192099.
XX 29-MAR-2000; 2000US-0193480.
XX 15-MAY-2000; 2000US-0205230.
XX 09-JUN-2000; 2000US-0211315.
```


PT treating acne vulgaris -

PS Claim 1; SEQ ID No 6; 1069pp; English.

XX Sequences AA59506-AAS59804 represent DNA molecules encoding

CC propionibacterium acnes immunogenic polypeptides. The proteins and their

CC associated DNA sequences are used in the treatment, prevention and

CC diagnosis of medical conditions caused by P. acnes. The disorders include

CC SAPHO syndrome (synovitis, acne, pustulosis, hypertonis and

CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved

CC in infections of bone, joints and the central nervous system, however it

CC is particularly involved in the inflammatory lesions associated with acne

CC vulgaris. A method for detecting the presence or absence of P. acnes in a

CC patient comprises contacting a sample with a binding agent that binds to

CC the proteins of the invention and determining the amount of bound protein

CC in the sample. The polypeptides may be used as antigens in the production

CC of antibodies specific for P. acnes proteins. These antibodies can be

CC used to downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the

CC polypeptides shown in AAU40231-AAU40470, AAU67466 and AAU67467.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 23474 BP; 4822 A; 7579 C; 6703 G; 4366 T; 4 other;

Query Match 51.0%; Score 20.4; DB 23; Length 23474;

Best Local Similarity 71.1%; Pred. No. 1.6e+02;

Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 atcactgaccctactctacctggctgactgggtgg 39

Db 1696 ATCTCTTACCGTGTCTCGTGGTGGTGGTGGTGG 1659

RESULT 12

ID AAK81665/c

XX AAK81665 standard; DNA; 24908 BP.

XX AAK81665;

XX

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36477.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226273.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259676.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure: SEQ ID NO 36477; 3071pp + Sequence Listing; English.

CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 24908 BP; 6781 A; 4773 C; 5094 G; 8260 T; 0 other;

Query Match 51.0%; Score 20.4; DB 22; Length 24908;
Best Local Similarity 71.1%; Pred No. 1.6e+02;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 atcaattgacctactctacctggctgctgctgggtgg 39
Db 16291 ATCACTTGAGCCCAAGATACAGAGGCTGCTGCTGAGTCG 16254
||||||| ||| ||| ||||| ||||| |||

RESULT 13
AAZ58767/c
ID AA258767 standard; DNA; 359 BP.
XX
AC AAZ58767;
XX
DT 17-APR-2000 (first entry)
XX
DE Human huntingtin-interacting protein (HtP) exon 15.
XX
KW Huntingtin Interacting Protein; HtP; death effector domain; DED; human;
KW apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;
KW Huntington's disease; neotrophic; anticonvulsant; cytostatic; exon;
KW chromosome 7q11.23; ds.

OS Homo sapiens.
XX
PN WO9960986-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11743.
XX
PR 27-MAY-1998; 98US-0085199.
XX
XX (UYBR-) UNIV BRITISH COLUMBIA.
PA (MERI) MERCK FROSST CANADA INC.
XX
PI Kalchman M, Hayden MR, Hackam A, Chopra V, Nicholson DW;
PI Vallaincourt JP, Rasper DM;
XX
DR WPI: 2000-097055/08.
XX
XX Novel proteins useful for treating Huntington's disease by gene therapy
XX techniques, and cancers -
XX
XX Example 10; Page 84; 91pp; English.

XX
XX The invention relates to Huntingtin Interacting Protein (HIP), that
XX includes a death effector domain (DED), suggesting apoptotic function.
XX Proteins with DED (referred as HIP-apoptosis modulating proteins) are
XX useful for inducing apoptotic death in cells. The HIP is a logical
XX target for therapy in Huntington's disease since it has a differential
XX apoptotic activity, modulated by interaction with Huntingtin having
XX normal and expanded repeats. HIP is also used as a therapeutic agent to
XX introduce apoptosis in cancer cells. Increasing expression of normal
XX (non-expanded) Huntingtin or the HIP-apoptotic modulator-binding
XX portion, a modified HIP-apoptotic modulator in which the DED has been
XX deleted, is helpful for treating Huntington's disease. Sequences
XX AAZ58754-780 represent human HIP1 exons 1 to 28 respectively.
XX
XX Sequence 359 BP; 82 A; 71 C; 127 G; 79 T; 0 other;

Query Match 50.5%; Score 20.2; DB 21; Length 359;

Query Match 50.5%; Score 20.2; DB 21; Length 359;
Best Local Similarity 75.8%; pred No 1e+02.

| | | |
|-----------------------|------------------|-------------------|
| Best local similarity | 75.8% | Fixed: NO. 1E+02; |
| Matches | 25; Conservative | 0; Mismatches |
| | | 8; Indels |
| | | 0; Gaps |

Qy 3 tcacttgaccctactctacctgggctggactgg 35

Db 188 TCCTCGTTCCAAATCTACCTGGGCTTGTCCTGG 156

RESULT 14

AAS61792/C

ID AAS61792 standard; cDNA; 601 BP.

AC AAS61792;

DT 29-JAN-2002 (first entry)

DE Lung small cell carcinoma antigen, cDNA #333.

KW Human; cytostatic; antitumour; lung small cell cancer antigen;

KW tumour; lung cancer; ss.

OS Homo sapiens.

PN WO200177168-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-US11859.

PR 11-APR-2000: 2000US-196780P.

PR 21-JUN-2000; 2000US-213361P.

PR 01-SEP-2000; 2000US-229763P..

PR 05-SEP-2000; 2000US-230629P.

PR 14-SEP-2000; 2000US-232565P.

PR 19-DEC-2000; 2000US-257037P.

PR 08-JAN-2001; 2001US-260796P.
v vXX
PA
/CORT-\ CORTYA CORP

| XX | PT | MT | Wage | Months | Cost | Cost | Cost | Cost | Cost |
|----|----|----|------|--------|------|------|------|------|------|
| XX | PT | MT | Wage | Months | Cost | Cost | Cost | Cost | Cost |

XX
DB WPT: 2002-010806/01

PT Lung tumour polynucleotide and polypeptides
PT diagnosis of cancer especially lung cancer

xx
PS
Claim 1: page 245. 29500. English

The invention relates to novel isolated lung small cell cancer antigen polynucleotides (I) and polypeptides (II) used in a method of detecting cancer in a patient. The method is optionally performed by utilising oligonucleotides (III), where the biological sample from the patient is contacted with (III), detecting the sample of polynucleotide hybridised to (III) in the sample and comparing the amount of polynucleotide to a predetermined cut-off value and thereby determining cancer in a patient. (I), (II) or antigen-presenting cells expressing (II) is useful for stimulating and/or expanding T cells specific for a tumour protein. The method comprises contacting T cells with one of the components under conditions to permit the stimulation and/or expansion of the cells. A composition comprising (I) is useful for stimulating an immune response in a patient and for inhibiting the development of a cancer especially lung cancer in a patient. An isolated T cell population is useful for removing tumour cells from the biological sample and for inhibiting the development of cancer in a patient. AAS61460-AAS61874 represent novel human lung small cell cancer antigen coding sequences of the invention.

SQ Sequence 601 BP; 183 A; 132 C; 185 G; 101 T; 0 other;

Query Match

Query Match
Best Local Similarity

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:46:31 ; Search time 2137.99 Seconds
(without alignments)
391.518 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 catcaactgaccctactctactctggtgactgggtggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgtg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| ----- | | | | | |

| | | | | | | | |
|---|----|------|------|--------|----|-----------|---------------------|
| C | 1 | 24.2 | 60.5 | 125817 | 2 | AC087325 | AC087325 Trypanoso |
| C | 2 | 23.8 | 59.5 | 40649 | 9 | AC003111 | AC003111 Human DNA |
| C | 3 | 23.8 | 59.5 | 42497 | 9 | AC004798 | AC004798 Homo sapi |
| C | 4 | 22.8 | 57.0 | 45420 | 2 | AC109521 | AC109521 Mus muscu |
| C | 5 | 22.8 | 57.0 | 157076 | 9 | AC019031 | AC019031 Homo sapi |
| C | 6 | 22.8 | 57.0 | 212671 | 2 | AC084849 | AC084849 Homo sapi |
| C | 7 | 22.8 | 57.0 | 213611 | 2 | AC091573 | AC091573 Homo sapi |
| C | 8 | 22.8 | 57.0 | 213621 | 9 | AF276759 | AF276759 Homo sapi |
| C | 9 | 22.6 | 56.5 | 112188 | 2 | AC098541 | AC098541 Rattus no |
| C | 10 | 22.6 | 56.5 | 162044 | 9 | AL390763 | AL390763 Human DNA |
| C | 11 | 22.6 | 56.5 | 163357 | 2 | AC023276 | AC023276 Homo sapi |
| C | 12 | 22.6 | 56.5 | 172385 | 2 | AL591127 | AL591127 Homo sapi |
| C | 13 | 22.6 | 56.5 | 178660 | 9 | HSG1150 | HSG1150 Homo sapi |
| C | 14 | 22.6 | 56.5 | 340000 | 9 | AP001700 | AP001700 Homo sapi |
| C | 15 | 22.4 | 56.0 | 333 | 11 | G42160 | G42160 ps701068 KW |
| C | 16 | 22.4 | 56.0 | 68118 | 2 | AC011113 | AC011113 Homo sapi |
| C | 17 | 22.4 | 56.0 | 139515 | 9 | HS38C16 | HS38C16 Human DNA |
| C | 18 | 22.4 | 56.0 | 150147 | 9 | AC002543 | AC002543 Homo sapi |
| C | 19 | 22.4 | 56.0 | 160543 | 9 | AC010271 | AC010271 Homo sapi |
| C | 20 | 22.4 | 56.0 | 164652 | 2 | AC021625 | AC021625 Homo sapi |
| C | 21 | 22.4 | 56.0 | 206717 | 2 | AL645535 | AL645535 Mus muscu |
| C | 22 | 22.4 | 56.0 | 222232 | 2 | AL645963 | AL645963 Mus muscu |
| C | 23 | 22.2 | 55.5 | 2544 | 9 | AK023107 | AK023107 Homo sapi |
| C | 24 | 22.2 | 55.5 | 142645 | 9 | AC008481 | AC008481 Homo sapi |
| C | 25 | 22.2 | 55.5 | 172705 | 9 | AC013272 | AC013272 Homo sapi |
| C | 26 | 22.2 | 55.5 | 176894 | 2 | AL355574 | AL355574 Homo sapi |
| C | 27 | 22.2 | 55.5 | 181890 | 2 | AL355987 | AL355987 Homo sapi |
| C | 28 | 22.2 | 55.5 | 187883 | 2 | AC008268 | AC008268 Homo sapi |
| C | 29 | 22.2 | 55.5 | 248846 | 2 | AC090008 | AC090008 Mus muscu |
| C | 30 | 22 | 55.0 | 767 | 8 | AF168146 | AF168146 Calliervo |
| C | 31 | 22 | 55.0 | 109255 | 2 | AC098163 | AC098163 Rattus no |
| C | 32 | 22 | 55.0 | 167118 | 2 | AC098129 | AC098129 Rattus no |
| C | 33 | 22 | 55.0 | 181807 | 9 | AL356267 | AL356267 Human DNA |
| C | 34 | 21.6 | 54.0 | 1511 | 8 | ZMANT2MU | X59086 Z. mays MANT |
| C | 35 | 21.6 | 54.0 | 2010 | 14 | AF109468 | AF109468 Feline ca |
| C | 36 | 21.6 | 54.0 | 104509 | 2 | AC095464 | AC095464 Rattus no |
| C | 37 | 21.6 | 54.0 | 110000 | 2 | AC103319 | AC103319 Rattus no |
| C | 38 | 21.6 | 54.0 | 115364 | 2 | AC094857 | AC094857 Rattus no |
| C | 39 | 21.6 | 54.0 | 173342 | 9 | AC023796 | AC023796 Homo sapi |
| C | 40 | 21.6 | 54.0 | 185306 | 2 | AC013570 | AC013570 Homo sapi |
| C | 41 | 21.6 | 54.0 | 209345 | 2 | AC106686 | AC106686 Homo sapi |
| C | 42 | 21.4 | 53.5 | 9685 | 1 | STY242516 | AJ242516 Salmonell |
| C | 43 | 21.4 | 53.5 | 22286 | 1 | AE008700 | AE008700 Salmonell |
| C | 44 | 21.4 | 53.5 | 75684 | 9 | AL512843 | AL512843 Human DNA |
| C | 45 | 21.4 | 53.5 | 119939 | 2 | AC098001 | AC098001 Rattus no |

ALIGNMENTS

| | |
|------------|--|
| RESULT | 1 |
| AC087325/c | |
| LOCUS | 125817 bp DNA linear HTG 17-JUL-2001 |
| DEFINITION | Trypanosoma brucei chromosome IV clone RPCI93-29M18, *** SEQUENCING IN PROGRESS ***, 1 ordered pieces. |
| ACCESSION | AC087325 |
| VERSION | AC087325.8 GI:14787198 |
| KEYWORDS | HTG: HTGS_PHASE2. |
| SOURCE | Trypanosoma brucei. |
| ORGANISM | Trypanosoma brucei. |
| REFERENCE | 1 (bases 1 to 125817) |
| AUTHORS | El-Sayed, N.M., Ghedin, E., Song, J., Larkin, C., Wanless, D., Jones, K., Peterson, J., Hou, L., Zhao, H., Mason, T., Millican, J., Pa, G., Van Aken, S., Uterback, T., Khalak, H.G., Gerard, C., Leech, V., Ullu, E., Melville, S., White, O., Adams, M.D., Donelson, J.E. and Fraser, C.M. |
| TITLE | Trypanosoma brucei GUTat10.1 RPCI93-29M18 BAC genomic sequence |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 125817) |
| AUTHORS | El-Sayed, N.M., Khalak, H. and Adams, M.D. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (25-DEC-2000) The Institute for Genomic Research, 9712 |


```

402/404 (99%).
(1800..2132) R34240 yh80f09.s1 Homo sapiens cDNA clone
136073 3' similar to contains TAR1 repetitive element;
Score: 614 Identity: 346/363 (95%).
(2249..1897) IN71602 yw40b10.r1 Homo sapiens cDNA clone
254683 5' similar to contains element MER86 repetitive
element; Score: 588 Identity: 335/357 (93%).
(2212..1967) AA304114 EST16821 Aorta endothelial cells,
TNF alpha-treated Homo sapiens cDNA 5' end; Score: 443
Identity: 239/255 (93%).
1947..1973
/rpt_family="POLY_A"
2096..2176
/rpt_family="(CAAAA)n"
complement(2286..2339)
/rpt_family="POLY_A"
2313..3073
/Note="DSS similarity to overlapping ESTs:
(2340..2685) AA458860 aa26405.s1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:814377 3'; Score: 683 Identity:
345/345 (100%).
(2787..2313) AA278877 zs79g12.r1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:703750 5' similar to contains
element MER22 repetitive element; Score: 930 Identity:
473/475 (99%).
(2341..2692) AA573302 nm52e03.s1 NCI_CGAP_Br2 Homo sapiens
cDNA clone IMAGE:1071868; Score: 684 Identity: 347/352
(98%).
(2342..2769) AA579808 nk43f03.s1 NCI_CGAP_GC2 Homo sapiens
cDNA clone IMAGE:1016285; Score: 848 Identity: 426/428
(99%).
(2345..2777) AA121152 z188e01.s1 Stratagene colon
(#937204) Homo sapiens cDNA clone 511704 3'; Score: 842
Identity: 430/433 (99%).
(3003..2614) AA121132 z188e01.r1 Stratagene colon
(#937204) Homo sapiens cDNA clone 511704 5'; Score: 726
Identity: 383/386 (99%).
(2496..2807) T07432 EST05321 Homo sapiens cDNA clone
HFBE195; Score: 365 Identity: 306/317 (96%).
(2950..2560) R68397 yh99h03.r1 Homo sapiens cDNA clone
137909 5'. Score: 687 Identity: 382/396 (96%).
(3073..2821) T85539 yd78h03.r1 Homo sapiens cDNA clone
114389 5'. Score: 463 Identity: 244/254 (96%).
(3035..2902) AA666963 vr86d06.s1 Knowles Solter mouse 2
cell Mus musculus cDNA clone 1135595 5'; (355..514) 80%
identity.
And many others.
2510..2589
/rpt_family="POLY_A"
complement(3776..3858)
/rpt_family="(CAG)n"
complement(4402..4458)
/Note="DSS similarity to AA666963 vr86d06.s1 Knowles
Solter mouse 2 cell Mus musculus cDNA clone 1135595 5';
(298..354) 88% identity.
complement(4595..4757)
/Note="DSS similarity to AA666963 vr86d06.s1 Knowles
Solter mouse 2 cell Mus musculus cDNA clone 1135595 5';
(135..297) 86% identity.
complement(4846..4973)
/Note="DSS similarity to AA666963 vr86d06.s1 Knowles
Solter mouse 2 cell Mus musculus cDNA clone 1135595 5';
(1..134) 84% identity.
5268..5308
/rpt_family="MER5A"
complement(6747..7049)
/rpt_family="AlusX"
8022..8145
/rpt_family="(TGG)n"
complement(8151..8405)
/Note="DSS similarity to AA057523 zf56e09.r1 Soares retina
N2b4HR Homo sapiens cDNA clone 380968 5' similar to
contains element MSR1 repetitive element; Score: 434

```

```

repeat_region
repeat_region
repeat_region
misc_feature

```

```

Identity: 236/255 (92%)
9400..9540
/rpt_family="(TGG)n"
9540..9586
/rpt_family="(CAGA)n"
complement(10070..10174)
/rpt_family="MIR"
11994..13585
/Note="DSS similarity to overlapping ESTs:
(11994..12459) N49889 yv24b06.s1 Homo sapiens cDNA clone
243635 3'; Score: 893 Identity: 461/467 (98%).
(12541..12078) AA493873 nh03d10.s1 NCI_CGAP_Thy1 Homo
sapiens cDNA clone IMAGE:943219; Score: 809 Identity:
443/457 (96%).
(12197..12697) AA429349 zv50g01.s1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 757104 3'; Score: 978
Identity: 495/501 (98%).
(12197..12702) AA291401 zt38f02.s1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 724635 3'; Score: 976
Identity: 500/506 (98%).
(12197..12788) N45196 yz14a11.s1 Homo sapiens cDNA clone
283004 3'; Score: 1068 Identity: 573/584 (98%).
(12880..12348) AA625129 af70a09.r1 Soares NHMpu S1 Homo
sapiens cDNA clone 1047352 5'; Score: 1017 Identity:
527/528 (99%).
(13237..12768) AA291565 zt38f02.r1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 724635 5'; Score: 895
Identity: 466/469 (99%).
(13237..12850) AA428761 zv50g01.r1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 757104 5'; Score: 776
Identity: 388/388 (100%). -Other overlapping EST matches:
T51194"
complement(13455..>38069)
/gene="HUNK1"
/product="H. sapiens HUNK1 (Y12059)"
complement(13455..>38069)
/gene="HUNK1"
complement(13585..14213)
/gene="HUNK1"

```

```

mrna

```

```

/gene="HUNK1"
/product="H. sapiens HUNK1 (Y12059)"
complement(13455..>38069)
/gene="HUNK1"
complement(13585..14213)
/gene="HUNK1"
/Note="DSS similarity to overlapping ESTs:
(13964..13585) R84483 y036f05.r1 Homo sapiens cDNA clone
180033 5'; Score: 668 Identity: 360/380 (94%).
(14132..13748) AA349685 EST56553 Infant brain Homo sapiens
cDNA 5' end; Score: 743 Identity: 380/386 (98%).
(14213..13751) H30841 y078c10.r1 Homo sapiens cDNA clone
184050 5'; Score: 885 Identity: 457/462 (98%).
(14213..13839) R88182 ym90f03.r1 Homo sapiens cDNA clone
166205 5'; Score: 724 Identity: 372/377 (98%).
(14226..14171) R29034 F0-114D 22 week old human fetal
liver cDNA library Homo sapiens cDNA clone F0-114D 5';
(56..1) 100% identity.
And many others.
16375..16403
/rpt_family="POLY_G"
17570..18018

```

```

misc_feature

```

```

/Note="DSS similarity to AA262797 zs24f07.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:686149 5'; Score: 833
Identity: 435/446 (97%).
complement(18294..18623)

```

```

repeat_region

```

```

misc_feature

```

```

misc_feature

```

```

Query Match 59.5%; Score 23.8; DB 9; Length 42497;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Oy 3 tcacttgacctactctacgtggctgacctgggt 37
||||| i | ||||| ||||| ||||| |||||

```

```

Db 36579 TCATGCAACCAATCCTAGCTGGCTGCACAGGT 36545

```

```

RESULT 4

```

```

AC109621/c AC109621
LOCUS Mus musculus clone RP23-330J17, LOW-PASS SEQUENCE SAMPLING.
DEFINITION Mus musculus clone RP23-330J17, linear HTG 06-FEB-2002
45420 bp DNA

```

```

ACCESSION      AC109621
VERSION        AC109621.1  GI:18543039
KEYWORDS       HTG: HTGS_PHASE0.
SOURCE         house mouse.
ORGANISM       Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus.
1 (bases 1 to 45420)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-330J17
Unpublished
2 (bases 1 to 45420)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepeil,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,Y., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zaloun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20876
Center clone name: 330_J_17
-----
* NOTE: This record contains 57 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 753: contig of 753 bp in length
* 754 853: gap of 100 bp
* 854 1553: contig of 700 bp in length
* 1554 1653: gap of 100 bp
* 1654 2333: contig of 680 bp in length
* 2334 2433: gap of 100 bp
* 2434 3164: contig of 731 bp in length
* 3165 3264: gap of 100 bp
* 3265 3967: contig of 703 bp in length
* 3968 4067: gap of 100 bp
* 4068 4772: contig of 705 bp in length
* 4773 4872: gap of 100 bp
* 4873 5531: contig of 659 bp in length
*
5532 5631: gap of 100 bp
5632 6291: contig of 660 bp in length
6292 6391: gap of 100 bp
6392 7098: contig of 707 bp in length
7099 7198: gap of 100 bp
7199 7883: contig of 685 bp in length
7884 7983: gap of 100 bp
7984 8697: contig of 714 bp in length
8698 8797: gap of 100 bp
8798 9506: contig of 709 bp in length
9507 9606: gap of 100 bp
9607 10317: contig of 711 bp in length
10318 10417: gap of 100 bp
10418 11109: contig of 692 bp in length
11110 11209: gap of 100 bp
11210 11918: contig of 709 bp in length
11919 12018: gap of 100 bp
12019 12709: contig of 691 bp in length
12710 12809: gap of 100 bp
12810 13509: contig of 700 bp in length
13510 13609: gap of 100 bp
13610 14299: contig of 690 bp in length
14300 14399: gap of 100 bp
14400 15105: contig of 706 bp in length
15106 15205: gap of 100 bp
15206 15908: contig of 703 bp in length
15909 16008: gap of 100 bp
16009 16713: contig of 705 bp in length
16714 16813: gap of 100 bp
16814 17522: contig of 709 bp in length
17523 17622: gap of 100 bp
17623 18299: contig of 677 bp in length
18300 18399: gap of 100 bp
18400 19099: contig of 700 bp in length
19100 19199: gap of 100 bp
19200 19903: contig of 704 bp in length
19904 20003: gap of 100 bp
20004 20699: contig of 696 bp in length
20700 20799: gap of 100 bp
20800 21508: contig of 709 bp in length
21509 21608: gap of 100 bp
21609 22316: contig of 708 bp in length
22317 22416: gap of 100 bp
22417 23121: contig of 705 bp in length
23122 23221: gap of 100 bp
23222 23888: contig of 667 bp in length
23889 23988: gap of 100 bp
23989 24689: contig of 701 bp in length
24690 24789: gap of 100 bp
24790 25491: contig of 702 bp in length
25492 25591: gap of 100 bp
25592 26254: contig of 663 bp in length
26255 26354: gap of 100 bp
26355 27058: contig of 704 bp in length
27059 27158: gap of 100 bp
27159 27861: contig of 703 bp in length
27862 27961: gap of 100 bp
27962 28666: contig of 705 bp in length
28667 28766: gap of 100 bp
28767 29476: contig of 710 bp in length
29477 29576: gap of 100 bp
29577 30272: contig of 696 bp in length
30273 30372: gap of 100 bp
30373 31062: contig of 690 bp in length
31063 31162: gap of 100 bp
31163 31867: contig of 705 bp in length
31868 31967: gap of 100 bp
31968 32670: contig of 703 bp in length
32671 32770: gap of 100 bp
32771 33475: contig of 705 bp in length
33476 33575: gap of 100 bp
33576 34289: contig of 714 bp in length
34290 34389: gap of 100 bp

```


Tirrell,A., Travers,M., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,W.M.
 Direct Submission
 Submitted (22-NOV-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBK

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: l1l002

Center clone name: 460_I-20

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 203142 bases at least Q40

Consensus quality: 208012 bases at least Q30

Consensus quality: 209766 bases at least Q20

Insert size: 200000; agarose-fp

Insert size: 210871; sum-of-contigs

Quality coverage: 5.9 in Q20 bases; agarose-fp

Quality coverage: 5.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 19 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * As soon as it is available and the accession number will
 * be preserved.

1 40594: contig of 40594 bp in length
 40595 40694: gap of 100 bp
 40695 41347: contig of 653 bp in length
 41348 41447: gap of 100 bp
 41448 42786: contig of 1339 bp in length
 42787 42886: gap of 100 bp
 42887 43895: contig of 1009 bp in length
 43896 43995: gap of 100 bp
 43996 45280: contig of 1285 bp in length
 45281 45380: gap of 100 bp
 45381 46855: contig of 1475 bp in length
 46856 46955: gap of 100 bp
 46956 48506: contig of 1551 bp in length
 48507 48606: gap of 100 bp
 48607 52429: contig of 3823 bp in length
 52430 52529: gap of 100 bp
 52530 54861: contig of 2332 bp in length
 54862 54961: gap of 100 bp
 54962 57521: contig of 2560 bp in length
 57522 57621: gap of 100 bp
 57622 62134: contig of 4513 bp in length
 62135 62234: gap of 100 bp
 62235 69724: contig of 7490 bp in length
 69725 69824: gap of 100 bp
 69825 77605: contig of 7781 bp in length
 77606 77705: gap of 100 bp
 77706 88690: contig of 10985 bp in length
 88691 88790: gap of 100 bp
 88791 104044: contig of 15254 bp in length
 104045 104144: gap of 100 bp
 104145 126880: contig of 22736 bp in length
 126881 126980: gap of 100 bp
 126981 163035: contig of 36055 bp in length
 163036 163135: gap of 100 bp
 163136 212019: contig of 48884 bp in length
 212020 212119: gap of 100 bp

* 212120 212671: contig of 552 bp in length.
 Location/Qualifiers

source

1..212671

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-460I20"

/clone.lib="RPC1-11 Human Male BAC"

1..40594

/note="assembly_fragment"

clone_end:Sp6

vector_side:left

40695..41347

/note="assembly_fragment"

41448..42786

/note="assembly_fragment"

42887..43895

/note="assembly_fragment"

43996..45280

/note="assembly_fragment"

45381..46855

/note="assembly_fragment"

46956..48506

/note="assembly_fragment"

48607..52429

/note="assembly_fragment"

52530..54861

/note="assembly_fragment"

54962..57521

/note="assembly_fragment"

57622..62134

/note="assembly_fragment"

62235..69724

/note="assembly_fragment"

69825..77605

/note="assembly_fragment"

77706..88690

/note="assembly_fragment"

88791..104044

/note="assembly_fragment"

104145..126880

/note="assembly_fragment"

126981..163035

/note="assembly_fragment"

163136..212019

/note="assembly_fragment"

212120..212671

/note="assembly_fragment"

clone_end:T7

vector_side:right

BASE COUNT 56524 a 48980 c 50176 g 55186 t 1805 others

ORIGIN

Query Match 57.0%; Score 22.8; DB 2; Length 212671;

Best Local Similarity 79.4%; Pred. No. 28;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 6 cttgacctactctacctggctgactggctggctgg 39

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 71803 CTTGGCCCTACTCTGCTTGGCAGGAGTGAGGG 71770

RESULT 7

AC091573/c

LOCUS AC091573

DEFINITION Homo sapiens chromosome 8 clone RP11-420F14 map 8, *** SEQUENCING

IN PROGRESS ***, 3 unordered pieces.

ACCESSION AC091573

VERSION AC091573.7 GI:18464127

KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE human.

213611 bp DNA linear HTG 11-FEB-2002


```

/note="Genefinder prediction"
1984. .2091
/note="Genefinder prediction"
complement(2020. .2135)
/note="XPOUND prediction, score = 0.379"
2488. .2669
/note="homology = 79.1%, counts = 13"
/rpt_family="atacacacacac repeat"
/rpt_type=TANDEM
2506. .2600
/note="AC repeat"
2530. .2582
/note="IR1, 79% complementary to IR1' (3707. .3759)"
/rpt_type=INVERTED
3227. .3243
/note="AT repeat"
complement(3248. .3302)
/note="match: GSS AQ213585"
3256. .3302
/note="match: GSS AQ260331"
complement(3435. .3561)
/note="84% identity: matches 134. .260 of consensus"
/rpt_family="AluJb"
3440. .3537
/note="Genefinder prediction"
complement(3613. .3672)
/note="match: GSS AQ234268"
3694. .3833
/note="homology = 60.0%, counts = 70"
/rpt_family="at repeat"
/rpt_type=TANDEM
3707. .3759
/note="IR1', 79% complementary to IR1 (2530. .2582)"
/rpt_type=INVERTED
4075. .4522
/note="match: EST AI859260"
4101. .4137
/note="WZEF prediction, score = 0.907"
complement(4698. .4711)
/note="XPOUND prediction, score = 0.816"
complement(5780. .5943)
/note="WZEF prediction, score = 0.811"
complement(5796. .5946)
/note="86% identity: matches 245. .395 of consensus"
/rpt_family="LI"
5796. .5946
/note="90% identity: matches 130. .280 of consensus"
/rpt_family="AluSb"
complement(5822. .5919)
/note="Genefinder prediction"
5950. .6001
/note="homology = 84.6%, counts = 4"
/rpt_family="aaaaaaaaaaat repeat"
/rpt_type=TANDEM
6683. .6695
/note="TC repeat"
complement((7117.7208). .7546)
/note="match: GSS B14745 B59625"
complement(7586. .7640)
/note="WZEF prediction, score = 0.568"
7791. .7803
/note="TC repeat"
complement(8072. .8088)
/note="XPOUND prediction, score = 0.241"
8090. .8103
/note="TGT repeat"
8178. .8217
/note="TTTA repeat"
8178. .8225
/note="homology = 95.8%, counts = 12"
/rpt_family="titta repeat"
/rpt_type=TANDEM
8267. .8450

```

```

repeat_region
/note="CpG_island (%GC=59.8, o/e=1.07, #CpGs=15)"
complement(8268. .8364)
/note="97% identity: matches 132. .228 of consensus"
/rpt_family="AluSb"
8273. .8364
/note="91% identity: matches 302. .393 of consensus"
/rpt_family="LI"
complement(8274. .8463)
/note="Genefinder prediction"
complement(8381. .8497)
/note="95% identity: matches 1. .117 of consensus"
/rpt_family="AluSb"
8381. .8486
/note="95% identity: matches 410. .515 of consensus"
/rpt_family="LI"
complement(8399. .8435)
/note="XPOUND prediction, score = 0.347"
complement(8860. .8991)
/note="GRAIL, score = 97%, comment = excellent"
complement(8873. .8934)
/note="XPOUND prediction, score = 0.391"
complement(8874. .8947)
/note="90% identity: matches 310. .383 of consensus"
/rpt_family="LI"
8972. .9037
/note="87% identity: matches 3445. .3510 of consensus"
/rpt_family="LI"
complement((9043.9056). (9084.9089))
/note="match: GSS AQ394050 AQ114719 AQ505132 AQ745232
AQ417781 AQ630059 AQ100697 B75664 AQ486114 AQ531167"
complement(9044. .9086)
/note="match: EST AI810082"
(9045.9047). (9086.9089)
/note="match: GSS AQ538467 B75590 AQ584427 AQ348442

Query Match 56.5%; Score 22.6; DB 9; Length 178660;
Best Local Similarity 75.7%; Pred. No. 35;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 tcacttgaccctactctacctggctggactgggtg9 39
||| | ||| ||||| ||||| |||||
Db 125893 TCCTTAAACTAGTCTACCTAGCTGGGCTGGTGG 125929

RESULT 14
AP001700 340000 bp DNA linear PRI 30-MAY-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 44/105.
DEFINITION AP001700 AL163245 BA000005
ACCESSION AP001700.1 GI:7768710
VERSION
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., and Schudy,A.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
20289799
2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,

```

Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Barand,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieseimann,L., Dagand,E., Wehrmaeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Vaspou,M.L.

TITLE JOURNAL

Direct Submission
Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)

COMMENT

On May 30, 2000 this sequence version replaced gi:7717301.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagami-hara 228-8555, Japan,
* e-mail: hattori@gsr.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and

* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e-mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and

* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
* e-mail: nshimizudmb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and

* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and

* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e-mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163245: Submitted (10-Apr-2000).

FEATURES

source

1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="KB851D4, 5' partial"
/clone_lib="Keio BAC library"
263. 627
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
920. 1710
/note="L1PB1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1720. 1752
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1753. 1834
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1835. 2193
/note="L1PB1"
/rpt_family="LINE/L1"

source

1. 104356
/note="Accession No. AP001607"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="KB851D4, 5' partial"
/clone_lib="Keio BAC library"
263. 627
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
920. 1710
/note="L1PB1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1720. 1752
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1753. 1834
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1835. 2193
/note="L1PB1"
/rpt_family="LINE/L1"

repeat_region

1. 104356
/note="Accession No. AP001607"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="KB851D4, 5' partial"
/clone_lib="Keio BAC library"
263. 627
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
920. 1710
/note="L1PB1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1720. 1752
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1753. 1834
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1835. 2193
/note="L1PB1"
/rpt_family="LINE/L1"

repeat_region

1. 104356
/note="Accession No. AP001607"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="KB851D4, 5' partial"
/clone_lib="Keio BAC library"
263. 627
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
920. 1710
/note="L1PB1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1720. 1752
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1753. 1834
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1835. 2193
/note="L1PB1"
/rpt_family="LINE/L1"

repeat_region

1. 104356
/note="Accession No. AP001607"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="KB851D4, 5' partial"
/clone_lib="Keio BAC library"
263. 627
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
920. 1710
/note="L1PB1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1720. 1752
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1753. 1834
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1835. 2193
/note="L1PB1"
/rpt_family="LINE/L1"

repeat_region
/rpt_type=DISPERSED
2209. 2521
/note="L1P"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2535. 2818
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(2819. 3122)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(3988. 4066)
/note="MER115"
/rpt_family="DNA"
/rpt_type=DISPERSED
complement(4100. 4253)
/note="MER5B"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
5543. 6453
/note="L1MA5A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
6798. 6847
/note="L1MA4A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(6954. 7127)
/note="MER20"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
7730. 7957
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
8282. 8343
/note="L1PA3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8503. 8813)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8875. 9276)
/note="L1ME1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
9581. 9944
/note="MLT1-INTERNAL"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
10088. 10623
/note="MLT1-INTERNAL"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
10699. 11123
/note="MLT1D"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
11589. 11659
/note="MLT1D"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
11661. 11847
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
12059. 12101
/note="(CAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM

```

Template: 16 ng Human genomic DNA
Primer: each 0.8 uM
dNTPs: each 200 uM
Amplifaq Polymerase: 2 units
Total Vol: 40 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3.

FEATURES
    source
        Location/Qualifiers
            1..333
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="KWOK"
            1..333
            1..22
                primer_bind
                primer_bind complement(312..333)
BASE COUNT      86 a   74 c   92 g   81 t
ORIGIN

Query Match      56.0%; Score 22.4; DB 11; Length 333;
Best Local Similarity 81.2%; Pred. No. 40;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      5  acttgaccctactctacctggctggactggg 36
        ||||| ||||| || ||||| ||||| |||
DB      132  ACTGTGTCCTCTGCACCTGGCTGGAGTGTG 163

Search completed: June 4, 2002, 17:47:45
Job time: 6733 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:08:54 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-13
Perfect score: 40
Sequence: 1 ctcgagccaccataaccctcaataactccaggattggg 40

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| C 1 | 24.6 | 61.5 | 296 | 9 | AW845395 | AW845395 CM4-CT004 |
| C 2 | 24.6 | 61.5 | 486 | 10 | BF740572 | BF740572 QV1-HB003 |
| C 3 | 23.8 | 59.5 | 250 | 10 | T08487 | T08487 EST06378 In |
| C 4 | 23.6 | 59.0 | 337 | 10 | T92497 | T92497 yz24b09.r1 |
| C 5 | 23 | 57.5 | 69 | 10 | BE878784 | BE878784 601493062 |
| C 6 | 23 | 57.5 | 118 | 9 | AW804230 | AW804230 PM3-UM008 |
| C 7 | 23 | 57.5 | 135 | 9 | BE005490 | BE005490 CM1-BN011 |
| C 8 | 23 | 57.5 | 147 | 9 | BE005492 | BE005492 CM1-BN011 |
| C 9 | 23 | 57.5 | 161 | 10 | N73434 | N73434 yz3le12.r1 |
| C 10 | 23 | 57.5 | 208 | 10 | BF090424 | BF090424 RCI-NT003 |
| C 11 | 23 | 57.5 | 223 | 9 | BE089733 | BE089733 QV0-BT070 |
| C 12 | 23 | 57.5 | 225 | 10 | BF340702 | BF340702 602035458 |
| C 13 | 23 | 57.5 | 227 | 9 | BE005486 | BE005486 CM1-BN011 |
| C 14 | 23 | 57.5 | 231 | 9 | AA327421 | AA327421 EST30756 |
| C 15 | 23 | 57.5 | 237 | 10 | BF037734 | BF037734 601462081 |
| C 16 | 23 | 57.5 | 243 | 10 | BG105201 | BG105201 602312948 |
| C 17 | 23 | 57.5 | 249 | 9 | AW804233 | AW804233 PM3-UM008 |

| | | | | | |
|------|------|-----|----|----------|--------------------|
| C 18 | 57.5 | 251 | 10 | BE622136 | BE622136 601440862 |
| C 19 | 57.5 | 252 | 10 | BG388188 | BG388188 602383704 |
| C 20 | 57.5 | 252 | 10 | BE869246 | BE869246 601445061 |
| C 21 | 57.5 | 252 | 10 | BE894967 | BE894967 601436172 |
| C 22 | 57.5 | 260 | 9 | AA101534 | AA101534 zn79c11.s |
| C 23 | 57.5 | 263 | 9 | AA383010 | AA383010 EST96488 |
| C 24 | 57.5 | 265 | 10 | BE878754 | BE878754 601493017 |
| C 25 | 57.5 | 273 | 10 | BF488202 | BF488202 QV0-EN005 |
| C 26 | 57.5 | 276 | 10 | BE617770 | BE617770 601441968 |
| C 27 | 57.5 | 277 | 10 | BG290247 | BG290247 602384948 |
| C 28 | 57.5 | 289 | 10 | T39447 | T39447 ya06a08.r1 |
| C 29 | 57.5 | 290 | 10 | BF740849 | BF740849 QV1-HB003 |
| C 30 | 57.5 | 299 | 9 | BE178852 | BE178852 PM1-HT060 |
| C 31 | 57.5 | 304 | 9 | AA384401 | AA384401 EST97914 |
| C 32 | 57.5 | 305 | 9 | AA853875 | AA853875 NHTBCae09 |
| C 33 | 57.5 | 307 | 9 | AW804241 | AW804241 PM3-UM008 |
| C 34 | 57.5 | 307 | 10 | T39458 | T39458 ya06b07.r2 |
| C 35 | 57.5 | 308 | 9 | AW804238 | AW804238 PM3-UM008 |
| C 36 | 57.5 | 308 | 9 | AA383479 | AA383479 EST96859 |
| C 37 | 57.5 | 313 | 10 | BE936437 | BE936437 RCI-NT003 |
| C 38 | 57.5 | 316 | 10 | BF342651 | BF342651 602013792 |
| C 39 | 57.5 | 317 | 10 | BG169657 | BG169657 602324628 |
| C 40 | 57.5 | 318 | 10 | BF968657 | BF968657 602270903 |
| C 41 | 57.5 | 321 | 10 | BI015712 | BI015712 PM0-ET020 |
| C 42 | 57.5 | 322 | 10 | BI013197 | BI013197 PM0-ET020 |
| C 43 | 57.5 | 322 | 10 | BI013200 | BI013200 PM0-ET020 |
| C 44 | 57.5 | 324 | 10 | BF884828 | BF884828 PM0-ET020 |
| C 45 | 57.5 | 326 | 9 | AA361993 | AA361993 EST71440 |

ALIGNMENTS

RESULT: 1
AW845395/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AW845395
CM4-CT0045-180200-512-a09 CT0045 Homo sapiens cDNA, mRNA sequence.
AW845395
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 296)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE
JOURNAL
MEDLINE
COMMENT
FEATURES
source

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=CM4-CT0045-180
200-512-a09&t3=2000-02-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 46
High quality sequence stop: 200.
Location/Qualifiers
i. .296
/organism="Homo sapiens"

```

BASE COUNT      129 a   129 c   110 g   118 t
ORIGIN

Query Match          61.5%;   Score 24.6;   DB 10;   Length 486;
Best Local Similarity 76.9%;   Pred. No. 35;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      2 tcagagccarcccataaacctcctaatactccaggagattggg 40
      ||| ||||| | ||| ||| ||| ||| ||| ||| ||| |||
Db      167 TCAGCGCACATTGAAGCCACCAAAACTGCAGGGATTGGG 205

```

| | |
|------------|--|
| RESULT | 3 |
| T08487 | |
| LOCUS | T08487 |
| DEFINITION | EST06378 Infant Brain, Bento Soares Homo sapiens cDNA clone HIB8E96 5' end, mRNA sequence. |
| | 250 bp mRNA linear EST 03-AUG-1993 |

| | |
|-----------|---|
| ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. |
| REFERENCE | 1 (bases 1 to 250) |
| AUTHORS | Adams,N.D., Soares,M.B., Kurlavage,A.R., Fields,C. and Venter,J.C. |
| TITLE | Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library |
| JOURNAL | Nature Genet. 4, 373-380 (1993) |
| MEDLINE | 94004965 |

CONTACT: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13 Reverse.
Location/Qualifiers
1..250

| BASE COUNT | 56 a | 84 c | 72 g | 31 t | 7 others |
|-----------------------|------|------|--------------|---------------|---------------------------------|
| ORIGIN | | | | | |
| Query Match | | | 59.5% | Score 23.8; | DB 10; Length 250; |
| Best Local Similarity | | | 77.8%; | Pred. No. 60; | |
| Matches 28; | | | Conservative | 0; | Mismatches 8; Indels 0; Gaps 0; |

| | | |
|------------|--|--|
| RESULT | 4 | |
| T92497/c | | |
| LOCUS | | |
| DEFINITION | | |
| ACCESSION | T92497 | 337 bp mRNA linear EST 22-MAR-1995 |
| VERSION | T92497.1 | ye24b03.r1 Stragatene lung (#937210) Homo sapiens cDNA clone |
| KEYWORDS | | IMAGE:118649 5', mRNA sequence. |
| EST. | | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| REFERENCE | 1 (bases 1 to 337) | |
| AUTHORS | Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., | |
| | Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins | |


```

DEFINITION yz3le12.rl Soares_multiple_sclerosis_2NBHMSF Homo sapiens cDNA
clone IMAGE:284686 5', mRNA sequence.
ACCESSION N73434
VERSION N73434.1 GI:1230719
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 161)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,
Underwood,K., Wohlmann,P., Wilson,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
TITLE 97044478
JOURNAL
MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1021 Std Error: 0.00
Seq primer: reverse ET
High quality sequence stop: 145.
location/Qualifiers
1..161
/organism="Homo sapiens"
/db_xref="GDB:3903214"
/db_xref="taxon:9606"
/clone_lib="IMAGE:284686"
/clone_lib="Soares_multiple_sclerosis_2NBHMSF"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT7T3D (Pharmacia) with a modified
polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not.I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Patima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 36 a 45 c 44 g 36 t
ORIGIN

Query Match 57.5%; Score 23; DB 10; Length 161;
Best Local Similarity 74.4%; Pred. No. 1.le+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccaccataaccctcaatactccaggattggg 40
||| ||||| | ||| ||||| ||||| ||||| |||||
Db 91 TCACCCACACTGTGAGCCACAAACTGCAGGATTGGG 53

RESULT 10
BF090424
LOCUS BF090424
DEFINITION RC1-NT0033-110900-111-c09 NT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF090424
VERSION BF090424.1 GI:10896134

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 208)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL 20202663
MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC1-NT0033-110
900-111-c09&t3=2000-09-11&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 208.
location/Qualifiers
1..208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0033"
/dev_stage="Adult"
/notes="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESSES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 53 a 62 c 51 g 42 t
ORIGIN

Query Match 57.5%; Score 23; DB 10; Length 208;
Best Local Similarity 74.4%; Pred. No. 1.le+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccaccataaccctcaatactccaggattggg 40
||| ||||| | ||| ||||| ||||| ||||| |||||
Db 115 TCAGCCACACTGTGAGCCACAAACTGCAGGATTGTG 153

RESULT 11
BE089733/c
LOCUS BE089733
DEFINITION QV0-BT0704-120500-224-c08 BT0704 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE089733
VERSION BE089733.1 GI:8480168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 223)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

```

```

FEATURES
    source              Location/Qualifiers
1..225
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4183263"
    /clone_lib="NCI_CGAP_Brn64"
    /tissue_type="glioblastoma with EGFR amplification"
    /lab_host="DHI0B (T1 phage-resistant)"
    /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
    Average insert size 1.57 kb. Constructed by Life
    Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT          49 a   48 c   60 g   68 t
ORIGIN

Query Match              57.5%;   Score 23;   DB 10;   Length 225;
Best Local Similarity    74.4%;   Pred. NO. 1.2e+02;
Matches                  29;   Conservative 0;   Mismatches 10;   Indels 0;   Gaps 0

Qy  2  tcgagccaccccaataccctcaatactccaggattggg  40
      |||||  |||||  |||||  |||||  |||||  |||||
Db  192  TCCACCCACACTTGAGCCACACAAACTGCAGGGATTGGG  154

RESULT 13
BE005486
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
    human.
ORGANISM
    Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS
    Dias Neto,E., Garcia Correa,A., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
TITLE
    Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
JOURNAL
MEDLINE
COMMENT
    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
    20202663
    Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2704921
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=st2-cm1-BN0117-110
    400-183-e09st3-2000-04-11&t4=1)
    Seq primer: puc 18 forward
    High quality sequence start: 17
    High quality sequence stop: 227.
    Location/Qualifiers
    1..227
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BN0117"
    /dev_stage="Adult"
    /note="Organ: breast normal; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of

```

tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 54 a 69 c 54 g 50 t

ORIGIN

Query Match 57.5%; Score 23; DB 9; Length 227;

Best Local Similarity 74.4%; Pred. No. 1.2e+02;

Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccataccctcaactccaggattggg 40

Db 154 TCACCCACACTGTGAAGCCACAAACTGCAGGATTGGG 192

RESULT 14

AA327421/c

LOCUS AA327421 231 bp mRNA linear EST 20-APR-1997

DEFINITION TEGT, mRNA sequence.

ACCESSION AA327421

VERSION AA327421.1 GI:1979666

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 231)

AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White

, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,

, L.M., Fitchugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodek, A.,

, Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,

Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,

Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

, Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,

Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,

Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon

, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and

Venter, J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl.), 3-174 (1995)

96026280

CONTACT: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (http://www.tigr.org/tdb/hgi/hgi.html)

Seq primer: M13 Reverse

Location/Qualifiers

1..231

/organism="Homo sapiens"

/db_xref="ATCC (inhost):128102"

/db_xref="taxon:9606"

/clone_lib="Colon I"

/dev_stage="adult"

/note="Organ: colon; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI"

49 a 50 c 64 g 68 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 57.5%; Score 23; DB 9; Length 231;

Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccataccctcaactccaggattggg 40

Db 182 TCACCCACACTGTGAAGCCACAAACTGCAGGATTGGG 144

RESULT 15

BF037734/c

LOCUS BF037734

DEFINITION 601462081f1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3865269 5',

mrna sequence.

ACCESSION BF037734

VERSION BF037734.1 GI:10746117

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 237)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue procurement: DCTD/DTF

cDNA Library Arrayed by: Life Technologies, Inc.

cDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9608 row: b column: 22

High quality sequence stop: 237.

Location/Qualifiers

1..237

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3865269"

/clone_lib="NIH_MGC_66"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.8 kb. Library constructed by Life

Technologies."

BASE COUNT 53 a 48 c 62 g 74 t

ORIGIN

Query Match

Best Local Similarity 57.5%; Score 23; DB 10; Length 237;

Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccataccctcaactccaggattggg 40

Db 224 TCACCCACACTGTGAAGCCACAAACTGCAGGATTGGG 186

Search completed: June 4, 2002, 17:08:55

Job time: 4528 sec

THIS PAGE BLANK (USPTO)


```

RESULT      6
US-08-735-609-9/c
; Sequence 9, Application US/08735609
; Patent No. 5994132
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey
; Amalfitano, Andrew
; Hausert, Michael A
;

```

```

7
RESULT
US-09-315-372-9/c
; Sequence 9, Application US/09315372
; Patent No. 6057158
;
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amafilano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.3
; CURRENT APPLICATION DATA:

```

Qy 5 agcacccataaacctcaataactccagggttggg 40

Db 2326 AGCTGCCCTCACCTGGATCCACGAGGACAGGG 2291

RESULT 10

US-09-221-235-10
; Sequence 10, Application US/09221235
; Patent No. 6043040
; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,235
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-235-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 11

US-09-221-928-10
; Sequence 10, Application US/09221928
; Patent No. 6121030
; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,928
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-928-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 12

US-09-221-527-10
; Sequence 10, Application US/09221527
; Patent No. 6146832

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,527
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-527-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 13

US-09-221-236-10

; Sequence 10, Application US/09221236

; Patent No. 6146841

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,236
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-236-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 14

US-09-221-416-10

; Sequence 10, Application US/09221416

; Patent No. 6153417

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,416
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115

```
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-416-10
```

```
Query Match      48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
Qy 1 ctcgagccaccataaccctcaatactaccagggattggg 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 ctccagcattcccaaaagctcttaattctccataaaatggg 994
```

```
RESULT 15
US-09-221-245-10
; Sequence 10, Application US/09221245
; Patent No. 6180358
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,245
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: US 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-245-10
```

```
Query Match      48.0%; Score 19.2; DB 4; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
Qy 1 ctcgagccaccataaccctcaatactaccagggattggg 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 ctccagcattcccaaaagctcttaattctccataaaatggg 994
```

```
Search completed: June 4, 2002, 17:47:11
Job time: 6584 sec
```


PT herpesvirus infections e.g. primary or chronic infections with human
PT cytomegalovirus

XX Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 10 A; 15 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcgagccacccataaacctcaatactccaggattggg 40
DB 1 ctcgagccacccataaacctcaatactccaggattggg 40

RESULT 2
AAV59067/c
ID AAV59067 standard; cDNA; 2634 BP.
AC AAV59067;
XX
DT 02-FEB-1999 (first entry)
XX
DE Bax inhibitor BI-1 cDNA.
XX
KW Bax inhibitor; BI-1; human; apoptosis; ss.
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FT CDS 73..786
FT /*tag= a

XX WO9840397-A1.
XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US05015.
XX 14-MAR-1997; 97US-0818514.

XX (BURN-) BURNHAM INST.

XX Reed JC, Xu Q;

XX WPI; 1998-531519/45.
XX P-PSDB; AAW73136.

XX Bax inhibitor proteins, BI-1 and BI-2 - useful e.g. to modulate
PT cellular apoptotic activity or identify agents altering BI-1 or BI-2
PT binding which can modulate apoptotic activity

XX Claim 2; Page 61-63; 80pp; English.

XX This cDNA clone codes for an inhibitor protein, termed BI-1 (see
CC AAW73136), of the pro-apoptotic protein Bax. Nucleic acids encoding
CC BI-1 and BI-2 (see AAV59068) were identified by suppression of
CC Bax-induced death of yeast cells transformed to express human Bax.
CC A human HepG2 cDNA library was used for library screening. The
CC invention provides vectors, optionally expression or viral vectors,
CC containing BI nucleic acids, and host cells containing these
CC vectors. The nucleic acids encoding BI-1/BI-2 can be used to

CC increase expression of these proteins in cells, or antisense
CC molecules prepared from them used to decrease expression. In
CC these ways, cellular apoptotic activity may be modulated (claimed).
CC The nucleic acids and complementary sequences are also useful as
CC probes to detect BI-encoding nucleic acid molecules in samples.

XX Sequence 2634 BP; 632 A; 626 C; 597 G; 779 T; 0 other;

Query Match 57.5%; Score 23; DB 19; Length 2634;
Best Local Similarity 74.4%; Pred. No. 4.1;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccacccataaacctcaatactccaggattggg 40
DB 2056 TCCACCCACACTTGAAGCCACAAACTGCAGGGATTGGG 2018

RESULT 3
AAF18054/c
ID AAF18054 standard; DNA; 2922 BP.
XX
AC AAF18054;

XX 14-MAR-2001 (first entry)
XX
DT Lung cancer associated polynucleotide sequence SEQ ID 73.

DE Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnery;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.

PN WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.
XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.

PI Ruben SM;

XX WPI; 2000-587514/55.
DR P-PSDB; AAB58178.

XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -

XX Claim 1; Page 550-551; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

Qy 7 ccaccataaccctcaatactccaggaattgg 40

AAK20013/C
ID AAK20013 standard: DNA: 121 bp.

PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 25; SEQ ID No 20651; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX Sequence 121 BP; 46 A; 20 C; 28 G; 27 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 121;
Best Local Similarity 76.5%; Pred. No. 14;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 7 ccacccataaacctcaataactccaggattggg 40
DB 38 CCTTCTATAATCCTCATTACTCAAGGTAATGGG 5

RESULT 9
ABA59145/c
ID ABA59145 standard; DNA; 473 BP.
AC ABA59145;
XX
XX 01-FEB-2002 (first entry)
XX Human foetal liver single exon nucleic acid probe #7450.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX Claim 1; SEQ ID NO 7450; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 7 ccacccataaacctcaataactccaggattggg 40
DB 347 CCTTCTATAATCCTCATTACTCAAGGTAATGGG 314

RESULT 10
ABA27922/c
ID ABA27922 standard; DNA; 473 BP.
XX
AC ABA27922;
XX
XX 23-JAN-2002 (first entry)
XX Probe #6388 for gene expression analysis in human heart cell sample.
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00666.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 6388; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI: 2001-656860/75.
DR P-PSDB; ABB61989.
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
PS Claim 1: SEQ ID NO 12758; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3896 BP; 1147 A; 931 C; 883 G; 935 T; 0 other;

Query Match 52.0%; Score 20.8; DB 23; Length 3896;
Best Local Similarity 70.0%; Pred. No. 38;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ctccgagccaccataaccctcaatactccagggtggg 40
||||| ||||| || || ||||| |||||
Db 3333 ctccgagccaccataaccagcagcagcgtccacggtggg 3372

Search completed: June 4, 2002, 17:56:14
Job time: 7102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:45:41 ; Search time 2137.99 Seconds
(without alignments)
391.518 Million cell updates/sec

Title: US-09-721-543A-13
Perfect score: 40
Sequence: 1 ctgagccacccataaccctcaatactccaggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|----|----|-------------|
|------------|-------|-------|-------|--------|----|----|-------------|

RESULT 1
AK026605/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

AK026605 1653 bp mRNA linear PRI 29-SEP-2000
Homo sapiens cDNA: FLJ22952 fis, clone KAT09742.

AK026605.1 GI:10439494

oligo capping; fis (full insert sequence).

Homo sapiens signal-ring cell carcinoma cell_line:KATO III cDNA to mRNA, clone_lib:KAT clone:KAT09742.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (sites)
Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

TITLE

NEDO human cDNA sequencing project

REFERENCE

Unpublished (2000)

AUTHORS

2 (bases 1 to 1653)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE

Direct Submission

JOURNAL

Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio

ALIGNMENTS

| | | | | | | |
|------|------|------|--------|----|-------------|--------------------|
| C 1 | 23 | 57.5 | 1653 | 9 | AK026605 | AK026605 |
| C 2 | 23 | 57.5 | 2600 | 9 | HSTEGT | HSTEGT |
| C 3 | 23 | 57.5 | 2609 | 9 | BC000916 | BC000916 Homo sapi |
| C 4 | 23 | 57.5 | 2634 | 6 | AR058920 | AR058920 Sequence |
| C 5 | 23 | 57.5 | 2634 | 6 | AR058921 | AR058921 Sequence |
| C 6 | 23 | 57.5 | 2634 | 6 | AR112791 | AR112791 Sequence |
| C 7 | 23 | 57.5 | 2634 | 6 | AR112792 | AR112792 Sequence |
| C 8 | 23 | 57.5 | 197189 | 2 | AC019168 | AC019168 Homo sapi |
| C 9 | 23 | 57.5 | 225867 | 2 | AC084037 | AC084037 Homo sapi |
| C 10 | 23 | 57.5 | 225818 | 2 | AC090998 | AC090998 Homo sapi |
| C 11 | 22.8 | 57.0 | 171185 | 2 | AP002424 | AP002424 Homo sapi |
| C 12 | 22.8 | 57.0 | 177097 | 2 | AP001569 | AP001569 Homo sapi |
| C 13 | 22.8 | 57.0 | 179726 | 9 | AC007052 | AC007052 Homo sapi |
| C 14 | 22.8 | 57.0 | 182411 | 2 | AC090408 | AC090408 Homo sapi |
| C 15 | 22.8 | 57.0 | 188357 | 9 | AL159996 | AL159996 Human DNA |
| C 16 | 22.8 | 57.0 | 189329 | 2 | AC090230 | AC090230 Homo sapi |
| C 17 | 22.8 | 57.0 | 196869 | 2 | AC087535 | AC087535 Homo sapi |
| C 18 | 22.8 | 57.0 | 200774 | 2 | AP001592 | AP001592 Homo sapi |
| C 19 | 22.4 | 56.0 | 137845 | 2 | AC010567 | AC010567 Homo sapi |
| C 20 | 22.4 | 56.0 | 151514 | 9 | HS471M13 | 297198 Human DNA s |
| C 21 | 22 | 55.0 | 452 | 11 | G22123 | G22123 human STS W |
| C 22 | 22 | 55.0 | 34887 | 3 | TBR012199 | AJ012199 Trypanoso |
| C 23 | 22 | 55.0 | 76727 | 9 | HS821D11 | AL021453 Human DNA |
| C 24 | 22 | 55.0 | 112022 | 9 | HSJ9E11 | AJ009611 Homo sapi |
| C 25 | 22 | 55.0 | 177540 | 9 | AC006538 | AC006538 Homo sapi |
| C 26 | 21.8 | 54.5 | 145607 | 2 | AC105341 | AC105341 Homo sapi |
| C 27 | 21.8 | 54.5 | 152339 | 2 | AC083931 | AC083931 Homo sapi |
| C 28 | 21.8 | 54.5 | 154165 | 2 | AC087620 | AC087620 Homo sapi |
| C 29 | 21.8 | 54.5 | 154599 | 2 | AC016997 | AC016997 Homo sapi |
| C 30 | 21.8 | 54.5 | 157021 | 2 | AC020991 | AC020991 Homo sapi |
| C 31 | 21.6 | 54.0 | 36415 | 9 | AC093645 | AC093645 Homo sapi |
| C 32 | 21.6 | 54.0 | 37470 | 9 | CNS00YVD | AL096810 Homo sapi |
| C 33 | 21.6 | 54.0 | 104436 | 9 | HS611N7 | AL035663 Human DNA |
| C 34 | 21.6 | 54.0 | 110000 | 2 | LMFLCHR12_4 | Continuation (5 of |
| C 35 | 21.6 | 54.0 | 122952 | 2 | AL513483 | AL513483 Homo sapi |
| C 36 | 21.6 | 54.0 | 173855 | 2 | OSN00181 | AL662983 Oryza sat |
| C 37 | 21.6 | 54.0 | 210442 | 2 | AL445199 | AL445199 Homo sapi |
| C 38 | 21.6 | 54.0 | 226860 | 9 | AL359920 | AL359920 Human DNA |
| C 39 | 21.4 | 53.5 | 54525 | 2 | AC103846 | AC103846 Homo sapi |
| C 40 | 21.4 | 53.5 | 70993 | 2 | AC102199 | AC102199 Mus muscu |
| C 41 | 21.4 | 53.5 | 161820 | 2 | AC095091 | AC095091 Rattus no |
| C 42 | 21.4 | 53.5 | 217345 | 2 | AC104393 | AC104393 Homo sapi |
| C 43 | 21.2 | 53.0 | 65824 | 2 | AC099899 | AC099899 Mus muscu |
| C 44 | 21.2 | 53.0 | 70594 | 2 | AC015786 | AC015786 Homo sapi |
| C 45 | 21.2 | 53.0 | 116252 | 9 | AL445228 | AL445228 Human DNA |


```

BASE COUNT      RKLMLAMNEKDKKEKK"
ORIGIN          606 a   625 c   602 g   776 t

Query Match           57.5%; Score 23; DB 9; Length 2609;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy    2 tcgagccaccataacacctcaatctcagggtggg 40
||| |||| | ||| |||| | ||| |||| | ||| |||| 
Db    2049 TCCACCCACTTTGAGGCCAACAACTGCAGGGATTGGG 2011


RESULT         AR058920/c
LOCUS          AR058920               2634 bp DNA linear PAT 29-SEP-1999
DEFINITION     Sequence 1 from patent US 5837838.
ACCESSION      AR058920
VERSION        AR058920.1 GI:5984497
KEYWORDS       .
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 2634)
AUTHORS       Reed,J.C. and Xu,Q.
TITLE         Bax inhibitor proteins
JOURNAL       Patent: US 5837838-A 1 17-NOV-1998;
FEATURES             Location/Qualifiers
                source            1..2634
                                /organism="unknown"
BASE COUNT     632 a   626 c   597 g   779 t
ORIGIN

Query Match           57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy    2 tcgagccaccataaacctcaatctcagggtggg 40
||| |||| | ||| |||| | ||| |||| | ||| |||| 
Db    2056 TCCACCACACTTGAAGCCACAATACTGCAGGGATGGG 2018


RESULT         AR058921/c
LOCUS          AR058921               2634 bp DNA linear PAT 16-MAY-2001
DEFINITION     Sequence 1 from patent US 6130317.
ACCESSION      AR112791
VERSION        AR112791.1 GI:14092691
KEYWORDS       .
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 2634)
AUTHORS       Reed,J.C. and Xu,Q.
TITLE         Bax inhibitor proteins
JOURNAL       Patent: US 6130317-A 1 10-OCT-2000;
FEATURES             Location/Qualifiers
                source            1..2634
                                /organism="unknown"
BASE COUNT     632 a   626 c   597 g   779 t
ORIGIN

Query Match           57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy    2 tcgagccaccataaacctcaatctcagggtggg 40
||| |||| | ||| |||| | ||| |||| | ||| |||| 
Db    2056 TCCACCACACTTGAAGCCACAATACTGCAGGGATGGG 2018


RESULT         AR112792/c
LOCUS          AR112792               2634 bp DNA linear PAT 16-MAY-2001
DEFINITION     Sequence 2 from patent US 6130317.
ACCESSION      AR112792
VERSION        AR112792.1 GI:14092692
KEYWORDS       .
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 2634)
AUTHORS       Reed,J.C. and Xu,Q.
TITLE         Bax inhibitor proteins
JOURNAL       Patent: US 6130317-A 2 10-OCT-2000;
FEATURES             Location/Qualifiers
                source            1..2634
                                /organism="unknown"
BASE COUNT     779 a   597 c   626 g   632 t
ORIGIN

Query Match           57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy    2 tcgagccaccataaacctcaatctcagggtggg 40
||| |||| | ||| |||| | ||| |||| | ||| |||| 
Db    579 TCCACCACACTTGAAGCCACAATACTGCAGGGATGGG 617


RESULT         AC019168
LOCUS          AC019168              197189 bp DNA linear HTG 07-JUL-2000
DEFINITION     Homo sapiens chromosome 12 clone RP11-161E16, WORKING DRAFT
SEQUENCE       SEQUENCE, 24 unordered pieces.
ACCESSION      AC019168
VERSION        AC019168.3 GI:7230912
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 197189)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 197189)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7021812.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0161E16
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: plasmid; 8%
Chemistry: Dye-terminator Big Dye; 8% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180505 bases at least Q40
Consensus quality: 186042 bases at least Q30
Consensus quality: 189445 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 194889; sum-of-contigs
Quality coverage: 3.69 in Q20 bases; agarose-fp
Quality coverage: 3.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1451: contig of 1451 bp in length
* 1452 1551: gap of unknown length
* 1552 2906: contig of 1355 bp in length
* 2907 3006: gap of unknown length
* 3007 4304: contig of 1298 bp in length
* 4305 4404: gap of unknown length
* 4405 5621: contig of 1217 bp in length
* 5622 5721: gap of unknown length
* 5722 8099: contig of 2378 bp in length
* 8100 8199: gap of unknown length
* 8200 10800: contig of 2601 bp in length
* 10801 10900: gap of unknown length
* 10901 14200: contig of 3300 bp in length
* 14201 14300: gap of unknown length
* 14301 17571: contig of 3271 bp in length
* 17572 17671: gap of unknown length
* 17672 21856: contig of 4185 bp in length
* 21857 21957: gap of unknown length
* 21958 25748: contig of 3792 bp in length
* 25749 25848: gap of unknown length
* 25849 29849: contig of 3994 bp in length
* 29843 29943: gap of unknown length
* 29944 35421: contig of 5479 bp in length
* 35422 35521: gap of unknown length
* 35522 40974: contig of 5453 bp in length
* 40975 41074: gap of unknown length
* 41075 48591: contig of 7517 bp in length
* 48592 48692: gap of unknown length
* 48693 58168: contig of 9477 bp in length
* 58169 58269: gap of unknown length
* 58270 68032: contig of 9764 bp in length
* 68033 68132: gap of unknown length
* 68133 77463: contig of 9331 bp in length

* 77464 77563: gap of unknown length
* 77564 93029: contig of 15466 bp in length
* 93030 107954: gap of unknown length
* 107955 108054: contig of 14825 bp in length
* 108055 121842: contig of 13788 bp in length
* 121843 121943: gap of unknown length
* 121944 137362: contig of 15420 bp in length
* 137363 137463: gap of unknown length
* 137464 155524: contig of 18062 bp in length
* 155525 15625: gap of unknown length
* 15626 173416: contig of 17792 bp in length
* 173417 173517: gap of unknown length
* 173518 197189: contig of 23673 bp in length.
FEATURES
Location/Qualifiers
1..197189
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-161E16"
1..1451
/note="assembly_name:Contig3"
1552..2906
/note="assembly_name:Contig4"
3007..4304
/note="assembly_name:Contig5"
4405..5621
/note="assembly_name:Contig6"
5722..8099
/note="assembly_name:Contig7"
8200..10800
/note="assembly_name:Contig8"
10901..14200
/note="assembly_name:Contig9"
14301..17571
/note="assembly_name:Contig10"
17672..21856
/note="assembly_name:Contig11"
21957..25748
/note="assembly_name:Contig12"
25849..29842
/note="assembly_name:Contig13"
29943..35421
/note="assembly_name:Contig14"
35522..40974
/note="assembly_name:Contig15"
41075..48591
/note="assembly_name:Contig16"
48692..58168
/note="assembly_name:Contig17"
58269..68032
/note="assembly_name:Contig18"
68133..77463
/note="assembly_name:Contig19"
77564..93029
/note="assembly_name:Contig20"
93130..107954
/note="assembly_name:Contig21"
clone_end:SP6
vector_side:right
108055..121842
/note="assembly_name:Contig22"
121943..137362
/note="assembly_name:Contig23"
137463..155524
/note="assembly_name:Contig24"
155625..173416
/note="assembly_name:Contig25"
173517..197189
/note="assembly_name:Contig26"
BASE COUNT 51312 a 47093 c 47437 g 49028 t 2319 others
ORIGIN

| | | | | |
|--|--------|---|---------------|----------------|
| Query Match | 57.5%; | Score 23; | DB 2; | Length 197189; |
| Best Local Similarity | 74.4%; | Pred. No. 27; | | |
| Matches | 29; | Conservative | 0; Mismatches | 10; Indels |
| | | | | Gaps |
| | | | | 0; |
| QY | 2 | tccagccaccctaacctcaatcattccaggattggg | 40 | |
| | | | | |
| Db | 123952 | TCCACCCATCTTGAGCCCAACAACACTGCAGGGATTGGG | 123990 | |
| RESULT | 9 | | | |
| AC084037 | | | | |
| LOCUS | | 215867 bp | DNA | linear |
| DEFINITION | | Homo sapiens chromosome 12q clone RP11-382G3, WORKING DRAFT SEQUENCE, 7 unordered pieces. | | |
| ACCESSION | | AC084037 | | |
| VERSION | | AC084037.28 | GI:18449688 | |
| KEYWORDS | | HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. | | |
| SOURCE | | human. | | |
| ORGANISM | | Homo sapiens | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | | 1 (bases 1 to 215867) | | |
| AUTHORS | | Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantune,K.C., Arc,J.R., Ayele,M., Banks,T., Barblar,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,P., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burgett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y.F., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lied,C., Liu,J., Liu,W., Loulsegod,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwu,S., Oguh,M., Okwuonu,G., Oragunye,N., Owiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojias,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R. | | |
| TITLE | | Direct Submission | | |
| JOURNAL | | Unpublished | | |
| REFERENCE | | 2 (bases 1 to 215867) | | |
| AUTHORS | | Worley,K.C. | | |
| TITLE | | Direct Submission | | |
| JOURNAL | | Submitted (11-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | |
| COMMENT | | On Jan 31, 2002 this sequence version replaced gi:15809132. | | |
| | | - - - - - Genome Center | | |

ACCESSION

AP002424
 VERSION AP002424.1 GI:8131688
 KEYWORDS HTGS_PHASE1; HTGS_DRAFT
 SOURCE Homo sapiens DNA, clone:RP11-873L22.
 ORGANISM Homo sapiens

REFERENCE

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE

JOURNAL Published Only in DataBase (2000) In press

REFERENCE

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan [E-mail:hattori@gsc.riken.go.jp]
 URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
 Fax:81-42-778-9924)

COMMENT

----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft18
 Center clone name: RP11-873L22
 ----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 156977 bases at least Q40
 Consensus quality: 164110 bases at least Q30
 Consensus quality: 167421 bases at least Q20
 Insert size: 169485; sum-of-contigs
 Quality coverage: 4.80x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of
 18 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

```

1      24923 contig of 24923 bp in length
25024  47004 contig of 21981 bp in length
47105  67070 contig of 19966 bp in length
82858  95544 contig of 12687 bp in length
95645  109592 contig of 13948 bp in length
109693  120167 contig of 10475 bp in length
120268  129255 contig of 8988 bp in length
129356  137003 contig of 7648 bp in length
137104  144376 contig of 7273 bp in length
150760  150659 contig of 6183 bp in length
155997  160522 contig of 5137 bp in length
160623  163263 contig of 4526 bp in length
163364  163194 contig of 2641 bp in length
165295  167704 contig of 1831 bp in length
167805  169750 contig of 2410 bp in length
169851  171185 contig of 1946 bp in length
        171185 contig of 1335 bp in length.
```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

*      1      24923: contig of 24923 bp in length
*      24924 25023: gap of 100 bp
*      25024 47004: contig of 21981 bp in length
*      47005 47104: gap of 100 bp
*      47105 67070: contig of 19966 bp in length
*      67071 67170: gap of 100 bp
*      67171 82757: contig of 15587 bp in length
*      82758 82857: gap of 100 bp
*      82858 95544: contig of 12687 bp in length
*      95545 95644: gap of 100 bp
*      95645 109592: contig of 13948 bp in length
*      109593 109692: gap of 100 bp
*      109693 120167: contig of 10475 bp in length
*      120168 120267: gap of 100 bp
*      120268 129255: contig of 8988 bp in length
*      129256 129355: gap of 100 bp
*      129356 137003: contig of 7648 bp in length
*      137004 137103: gap of 100 bp
*      137104 144376: contig of 7273 bp in length
*      144377 144476: gap of 100 bp
*      144477 150659: contig of 6183 bp in length
*      150660 150759: gap of 100 bp
*      150760 155896: contig of 5137 bp in length
*      155897 155996: gap of 100 bp
*      155997 160522: contig of 4526 bp in length
*      160523 160622: gap of 100 bp
*      160623 163263: contig of 2641 bp in length
*      163264 163363: gap of 100 bp
*      163364 165194: contig of 1831 bp in length
*      165195 165294: gap of 100 bp
*      165295 167704: contig of 2410 bp in length
*      167705 167804: gap of 100 bp
*      167805 169750: contig of 1946 bp in length
*      169751 169850: gap of 100 bp
*      169851 171185: contig of 1335 bp in length.
```

FEATURES

```

Source
1. .171185
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="18"
   /map="18q21"
   /clone="RP11-873L22"

misc_feature
1. .24923
   /note="assembly_fragment"
   25024..47004
   /note="assembly_fragment"
   47105..67070
   /note="assembly_fragment"
   67171..82757
   /note="assembly_fragment"
   82858..95544
   /note="assembly_fragment"
   95645..109592
   /note="assembly_fragment"
   109693..120167
   /note="assembly_fragment"
   120268..129255
   /note="assembly_fragment"
   129356..137003
   /note="assembly_fragment clone_end:SP6 vector_side:right"
   137104..144376
   /note="assembly_fragment"
   144477..150659
   /note="assembly_fragment"
   150760..155896
   /note="assembly_fragment"
   155997..160522
   /note="assembly_fragment"
   160623..163263
   /note="assembly_fragment"
   163364..165194
   /note="assembly_fragment clone_end:T7 vector_side:left"
   165295..167704
```



```
repeat_region /map="18"
repeat_region /clone="hrpk_411_h_24"
repeat_region /clone_lib="RCI-11 human BAC library"
repeat_region 898..1014
repeat_region /rpt_family="L1MD3"
repeat_region complement(1035..1333)
repeat_region /rpt_family="AluSx"
repeat_region 2987..3030
repeat_region /rpt_family="CT-rich"
repeat_region 3087..3291
repeat_region /rpt_family="L1MC4"
repeat_region 4298..4330
repeat_region /rpt_family="(CA)n"
repeat_region 5344..5365
repeat_region /rpt_family="AT-rich"
repeat_region complement(5552..5724)
repeat_region /rpt_family="AluJ"
repeat_region 5731..11847
repeat_region /rpt_family="L1PA2"
repeat_region 11857..11891
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(11893..12002)
repeat_region /rpt_family="AluJ"
repeat_region 11925..11948
repeat_region /note="Single-stranded coverage."
repeat_region 12026..12321
repeat_region /rpt_family="AluSg"
repeat_region complement(12605..12846)
repeat_region /rpt_family="MIR"
repeat_region 13352..13660
repeat_region /rpt_family="AluSc"
repeat_region complement(13741..14128)
repeat_region /rpt_family="MLT1B"
repeat_region 14425..14456
repeat_region /rpt_family="AT-rich"
repeat_region 14457..14978
repeat_region /rpt_family="MLT1F"
repeat_region 15496..15834
repeat_region /rpt_family="THE1B"
repeat_region 16325..16628
repeat_region /rpt_family="AluYb8"
repeat_region 16753..17086
repeat_region /rpt_family="MER7A"
repeat_region 17543..17601
repeat_region /rpt_family="MIR"
repeat_region 18075..18442
repeat_region /rpt_family="MLT1B"
repeat_region complement(18735..18882)
repeat_region /rpt_family="MIR"
repeat_region complement(18948..19083)
repeat_region /rpt_family="L1MB8"
repeat_region complement(19190..19426)
repeat_region /rpt_family="AluJo"
repeat_region 19431..19750
repeat_region /rpt_family="L1MB3"
repeat_region 20434..20555
repeat_region /rpt_family="MLT1A1"
repeat_region 20556..20863
repeat_region /rpt_family="AluSx"
repeat_region 20864..21080
repeat_region /rpt_family="MLT1A1"
repeat_region 21528..21565
repeat_region /rpt_family="AT-rich"
repeat_region 21575..21668
repeat_region /rpt_family="L2"
repeat_region complement(21849..21991)
repeat_region /rpt_family="MER5B"
repeat_region 23147..23168
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(23169..23384)
repeat_region /rpt_family="AluJo"
repeat_region 23385..23419
repeat_region /rpt_family="AT-rich"
```

```
repeat_region complement(23799..23846)
repeat_region /rpt_family="L2"
repeat_region complement(23947..24060)
repeat_region /rpt_family="L2"
repeat_region 24066..24206
repeat_region /rpt_family="BC200"
repeat_region complement(24347..24514)
repeat_region /rpt_family="MER21B"
repeat_region 24782..25203
repeat_region /rpt_family="L1MB2"
repeat_region 25385..26005
repeat_region /rpt_family="L1MB2"
repeat_region 26035..26336
repeat_region /rpt_family="AluSx"
repeat_region 26432..26604
repeat_region /rpt_family="L1ME"
repeat_region 26610..26645
repeat_region /rpt_family="AT-rich"
repeat_region complement(27113..27422)
repeat_region /rpt_family="AluSp"
repeat_region 27865..28190
repeat_region /rpt_family="AluSg"
repeat_region 28250..28272
repeat_region /rpt_family="AT-rich"
repeat_region 28363..28496
repeat_region /rpt_family="HAL1"
repeat_region complement(28497..28796)
repeat_region /rpt_family="AluSg"
repeat_region 28797..28869
repeat_region /rpt_family="HAL1"
repeat_region 28993..29027
repeat_region /rpt_family="AT-rich"
repeat_region 29107..29134
repeat_region /rpt_family="(TATATG)n"
repeat_region 29151..29182
repeat_region /rpt_family="(TG)n"
repeat_region complement(29821..30164)
repeat_region /rpt_family="MER7A"
repeat_region complement(30236..30522)
repeat_region /rpt_family="AluJo"
repeat_region 30576..30723
repeat_region /rpt_family="MIR"
repeat_region 30927..31271
repeat_region /rpt_family="MLT1K"
repeat_region 32382..32437
repeat_region /rpt_family="(CA)n"
repeat_region 32480..32502
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(32503..32751)
repeat_region /rpt_family="AluSx"
repeat_region 34312..34476
repeat_region /rpt_family="MIR"
repeat_region complement(34648..34953)
repeat_region /rpt_family="AluSx"
repeat_region complement(35550..35947)
repeat_region /rpt_family="THE1B"
repeat_region 36048..36070
repeat_region /rpt_family="(TTTAA)n"
repeat_region complement(36072..36366)
repeat_region /rpt_family="AluJo"
repeat_region 36525..36587
repeat_region /rpt_family="AT-rich"
repeat_region 36792..36849
repeat_region /rpt_family="AT-rich"
repeat_region complement(36903..37092)
repeat_region /rpt_family="L1MA7"
repeat_region 37502..37812
```

Query Match Best Local Similarity 59.0%; Score 22.8; DB 9; Length 179726;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctcgagccacccataaccctcaataactccagg 34

Db 78718 CTCAGCCTCCCCATGCCCCCTCATTTCTCCAGTG 78685

* as soon as it is available and the accession number will
* be preserved

| | | | |
|---|--------|---------|------------------------------|
| * | 804 | 903: | contig of 803 bp in length |
| * | 804 | 903: | gap of 100 bp |
| * | 904 | 1787: | contig of 884 bp in length |
| * | 1788 | 1887: | gap of 100 bp |
| * | 1888 | 3003: | contig of 1116 bp in length |
| * | 3004 | 3103: | gap of 100 bp |
| * | 3104 | 5498: | contig of 2395 bp in length |
| * | 5499 | 5598: | gap of 100 bp |
| * | 5599 | 7587: | contig of 1989 bp in length |
| * | 7588 | 7687: | gap of 100 bp |
| * | 7688 | 9739: | contig of 2052 bp in length |
| * | 9740 | 9839: | gap of 100 bp |
| * | 9840 | 13254: | contig of 3415 bp in length |
| * | 13255 | 13354: | gap of 100 bp |
| * | 13355 | 17739: | contig of 4385 bp in length |
| * | 17740 | 17839: | gap of 100 bp |
| * | 17840 | 25230: | contig of 7391 bp in length |
| * | 25231 | 25330: | gap of 100 bp |
| * | 25331 | 32077: | contig of 6747 bp in length |
| * | 32078 | 32177: | gap of 100 bp |
| * | 32178 | 48716: | contig of 16539 bp in length |
| * | 48717 | 48816: | gap of 100 bp |
| * | 48817 | 68998: | contig of 20182 bp in length |
| * | 68999 | 69098: | gap of 100 bp |
| * | 69099 | 103010: | contig of 33912 bp in length |
| * | 103011 | 103110: | gap of 100 bp |
| * | 103111 | 182411: | contig of 79301 bp in length |

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .182411 |

[illegible]

| BASE COUNT | ORIGIN | 56444 a | 35140 c | 35375 g | 54140 t | 1312 others |
|--------------------------|--------|---------|---------|---------|---------|-------------|
| /note= assembly_fragment | | | | | | |

```
Query Match      57.0%; Score 22.8; DB 2; Length 182411;
Best Local Similarity 79.4%; Pred. NO. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0;
```

| | |
|------------|---|
| RESULT 14 | |
| AC090408/c | 182411 bp DNA linear HTG 09-MAY-2001 |
| LOCUS | Homo sapiens chromosome 18 clone RP11-859C21 map 18, WORKING DRAFT |
| DEFINITION | SEQUENCE, 14 unordered pieces. |
| AC090408 | |
| AC090408.2 | GI-13357356 |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| AUTHORS | Mamalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| TITLE | 1 (bases 1 to 182411) |
| JOURNAL | Birren,B., Linton,L., Nusbaum,C. and Lander,E. |
| REFERENCE | Homo sapiens chromosome 18, clone RP11-859C21 |
| AUTHORS | 2 (bases 1 to 182411) |
| | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., |
| | Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., |
| | Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., |
| | Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., |
| | Dodges,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., |
| | Cardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., |
| | Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., |
| | Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T., |
| | Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., |
| | Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., |
| | McPheeters,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., |
| | Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., |
| | O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., |
| | Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., |
| | Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., |
| | Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., |
| | Sougnuez,C., Spencer,B., Stange-Rhmann,N., Stojanovic,N., |
| | Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., |
| | Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., |
| | Willson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., |
| | Zembek,L., Zimmer,A. and Zody,M. |

TITLE Direct Submission
JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On war 16 2001 this sequence version replaced g1.1258044. All repeats were identified using RepeatMasker.
Smit, F.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

```

----- Project Information
Center project name: L12718
Center clone name: 859_C_21
----- Summary Statistics

```

Summary statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 177348 bases at least Q40
Consensus quality: 179869 bases at least Q30
Consensus quality: 180657 bases at least Q20
Insert size: 18111; sum-of-contigs
Quality coverage: 6.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

Job time: 6659 sec

```
repeat_region 30822..31276
/note="MIR repeat: matches 70. .532 of consensus"
repeat_region 31655..32020
/note="RHE1B repeat: matches 1. .364 of consensus"
misc_feature 32181..32674
/note="match: GSS: Em:AQ822292"
repeat_region 32279..32403
/note="MIR repeat: matches 57. .192 of consensus"
misc_feature complement(32857..33263)
/note="match: GSS: Em:AQ139531"
repeat_region 33279..33551
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 35349..35388
/note="10 copies 4 mer acac 92% conserved"
repeat_region 35353..35388
/note="3 copies 12 mer 94% conserved"
repeat_region 35759..36002
/note="MIR repeat: matches 2. .246 of consensus"
misc_feature 36104..36469
/note="match: GSS: Em:AQ031409"
repeat_region 37497..37545
/note="MIR repeat: matches 81. .127 of consensus"
repeat_region 37664..38248
/note="MIR2D repeat: matches 1. .553 of consensus"
repeat_region 38259..38373
/note="MIR repeat: matches 122. .249 of consensus"
repeat_region 38334..38392
/note="L2 repeat: matches 2692. .2750 of consensus"
misc_feature 38609..38887
/note="match: STS: Em:G13414"
repeat_region 38781..38991
/note="L1R33 repeat: matches 302. .518 of consensus"
repeat_region 39086..39212
/note="L1R33 repeat: matches 29. .165 of consensus"
repeat_region 39249..39483
/note="MER4D repeat: matches 1. .234 of consensus"
repeat_region 39483..40142
/note="MER4D repeat: matches 359. .1017 of consensus"
repeat_region 40325..40415
/note="MER94 repeat: matches 39. .134 of consensus"
repeat_region 41795..41902
/note="L1R33 repeat: matches 370. .472 of consensus"
repeat_region 41903..42192
/note="AluSg repeat: matches 1. .291 of consensus"
repeat_region 42193..42241
/note="L1R33 repeat: matches 472. .521 of consensus"
repeat_region 43494..43613
/note="L2 repeat: matches 2628. .2748 of consensus"
repeat_region 43695..43882
/note="FRAM repeat: matches 0. .164 of consensus"
repeat_region 44019..44082
/note="L2 repeat: matches 2356. .2428 of consensus"
repeat_region 44083..44384
/note="L1P4A repeat: matches 5841. .6144 of consensus"
repeat_region 44385..44529
/note="L2 repeat: matches 2252. .2356 of consensus"
repeat_region 44541..44628
/note="MIR repeat: matches 128. .229 of consensus"
repeat_region 44631..44747
/note="MER33 repeat: matches 203. .324 of consensus"
repeat_region 44779..44985

Query Match 57.0%; Score 22.8; DB 9; Length 188357;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 ccacccataacccataactccaggattggg 40
||||||| | ||||| ||||| |||
Db 56526 CCACCCATGCTCTCAAGACTCCAGGGAAGG 56559
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 15:53:27 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-12
Perfect score: 40
Sequence: 1 ctgagccaccataaccctcaatactccaggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_estbha:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estovl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| C 1 | 24.6 | 61.5 | 296 | 9 | AW845395 | AW845395 CM4-CT004 |
| C 2 | 24.6 | 61.5 | 486 | 10 | BF740572 | BF740572 QV1-HB003 |
| C 3 | 23.8 | 59.0 | 250 | 10 | T08487 | T08487 EST06378 In |
| C 4 | 23.6 | 59.0 | 337 | 10 | T92497 | T92497 ye24b09.r1 |
| C 5 | 23 | 57.5 | 69 | 10 | BE878784 | BE878784 601493062 |
| C 6 | 23 | 57.5 | 118 | 9 | AW804230 | AW804230 PM3-UM008 |
| C 7 | 23 | 57.5 | 135 | 9 | BE005490 | BE005490 CM1-BN011 |
| C 8 | 23 | 57.5 | 147 | 9 | BE005492 | BE005492 CM1-BN011 |
| C 9 | 23 | 57.5 | 161 | 10 | N73434 | N73434 yz31e12.r1 |
| C 10 | 23 | 57.5 | 208 | 10 | BF090424 | BF090424 RCL1-NT003 |
| C 11 | 23 | 57.5 | 223 | 9 | BE089733 | BE089733 QV0-BT070 |
| C 12 | 23 | 57.5 | 225 | 10 | BF340702 | BF340702 602035458 |
| C 13 | 23 | 57.5 | 227 | 9 | BE005486 | BE005486 CM1-BN011 |
| C 14 | 23 | 57.5 | 231 | 9 | AA327421 | AA327421 EST30756 |
| C 15 | 23 | 57.5 | 237 | 10 | BF037734 | BF037734 601462081 |
| C 16 | 23 | 57.5 | 243 | 10 | BE105201 | BE105201 602312948 |
| C 17 | 23 | 57.5 | 249 | 9 | AW804233 | AW804233 PM3-UM008 |

| | | | | | | |
|------|----|------|-----|----|-----------|---------------------|
| C 18 | 23 | 57.5 | 251 | 10 | BE622136 | BE622136 601440862 |
| C 19 | 23 | 57.5 | 252 | 10 | BE288188 | BE288188 602383704 |
| C 20 | 23 | 57.5 | 252 | 10 | BE869246 | BE869246 601445061 |
| C 21 | 23 | 57.5 | 252 | 10 | BE894967 | BE894967 601436172 |
| C 22 | 23 | 57.5 | 260 | 9 | AA101534 | AA101534 zn79c11.s |
| C 23 | 23 | 57.5 | 263 | 9 | AA383010 | AA383010 EST96488 |
| C 24 | 23 | 57.5 | 265 | 10 | BE878754 | BE878754 601493017 |
| C 25 | 23 | 57.5 | 273 | 10 | BF484202 | BF484202 QV0-EN005 |
| C 26 | 23 | 57.5 | 276 | 10 | BE617770 | BE617770 601441968 |
| C 27 | 23 | 57.5 | 277 | 10 | BE6290247 | BE6290247 602384948 |
| C 28 | 23 | 57.5 | 289 | 10 | T39447 | T39447 ya06a08.r1 |
| C 29 | 23 | 57.5 | 290 | 10 | BF740849 | BF740849 QV1-HB003 |
| C 30 | 23 | 57.5 | 299 | 9 | BE178852 | BE178852 PM1-HT060 |
| C 31 | 23 | 57.5 | 304 | 9 | AA384401 | AA384401 EST97914 |
| C 32 | 23 | 57.5 | 305 | 9 | AA853875 | AA853875 NHTBCae09 |
| C 33 | 23 | 57.5 | 307 | 9 | AW804241 | AW804241 PM3-UM008 |
| C 34 | 23 | 57.5 | 307 | 10 | T39458 | T39458 ya06b07.r2 |
| C 35 | 23 | 57.5 | 308 | 9 | AW804238 | AW804238 PM3-UM008 |
| C 36 | 23 | 57.5 | 308 | 9 | AA383479 | AA383479 EST96859 |
| C 37 | 23 | 57.5 | 313 | 10 | BE936437 | BE936437 RCL1-NT003 |
| C 38 | 23 | 57.5 | 316 | 10 | BF342651 | BF342651 602013792 |
| C 39 | 23 | 57.5 | 317 | 10 | BG169657 | BG169657 602324628 |
| C 40 | 23 | 57.5 | 318 | 10 | BF968657 | BF968657 602270903 |
| C 41 | 23 | 57.5 | 321 | 10 | BI015712 | BI015712 PM0-ET020 |
| C 42 | 23 | 57.5 | 322 | 10 | BI013197 | BI013197 PM0-ET020 |
| C 43 | 23 | 57.5 | 322 | 10 | BI013200 | BI013200 PM0-ET020 |
| C 44 | 23 | 57.5 | 324 | 10 | BF884828 | BF884828 PM0-ET020 |
| C 45 | 23 | 57.5 | 326 | 9 | AA361993 | AA361993 EST71440 |

ALIGNMENTS

RESULT 1
AW845395/c
LOCUS 296 bp mRNA linear EST 19-MAY-2000
CM4-CT0045-180200-512-a09 CT0045 Homo sapiens cDNA, mRNA sequence.
AW845395
VERSION AW845395.1 GI:7940912
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 296)
AUTHORS Nagai, M.A., da Silva W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Briones, M.R., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-cm4-CT0045-180200-512-a09st3-2000-02-18st4=1)
Seq primer: puc 18 forward
High quality sequence start: 46
High quality sequence stop: 200.
Location/Qualifiers
1. .296
/organism="Homo sapiens"

| | | | | |
|--------------------------|---|-------|-------|-------|
| BASE COUNT | 129 a | 129 c | 110 g | 118 t |
| ORIGIN | | | | |
| Query Match | 61.5%; Score 24.6; DB 10; Length 486; | | | |
| Best Local Similarity | 76.9%; Pred. NO. 35; | | | |
| Matches 30; Conservative | 0; Mismatches 9; Indels 0; Gaps 0; | | | |
| Qy | 2 tcagagccaccccataaccctcaataactccaggagtggg 40 | | | |
| | | | | |
| Db | 167 TCAGGCACACTTGAGCCCAAAACTGCAGGGATTGGG 205 | | | |

| | |
|------------|---|
| RESULT | 3 |
| T08487 | |
| LOCUS | T08487 |
| DEFINITION | EST06378 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBB296 5' end, mRNA sequence. |
| | 250 bp |
| | mRNA |
| | linear |
| | EST 03-AUG-1993 |

| | |
|-----------|--|
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. I. (bases 1 to 250) |
| AUTHORS | Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C. |
| TITLE | Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library |
| JOURNAL | Nature Genet. 4, 373-380 (1993) |
| PMID | 94004905 |

CONTACT: ADAMS, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdamads@tigr.org
Seq primer: M13 Reverse.
Location/Qualifiers
1..250
FEATURES
source

| | | | | | |
|-----------------------|---|------|------|------|----------|
| BASE COUNT | 56 a | 84 c | 72 g | 31 t | 7 others |
| ORIGIN | /clone_id= infant brain; Bento Soares | | | | |
| Query Match | 59.5%; Score 23.8; DB 10; Length 250; | | | | |
| Best Local Similarity | 77.8%; Pred. No. 60; | | | | |
| Matches 28; | Conservative 0; Mismatches 8; Indels 0; Gaps 0; | | | | |

| | | |
|------------|---|--|
| RESULT | 4 | |
| T92497/c | | |
| LOCUS | | |
| DEFINITION | | |
| ACCESSION | T92497 | 337 bp mRNA linear EST 22-MAR-1995 |
| VERSION | T92497.1 | ye24b09.r1 Stratagene lung (#937210) Homo sapiens cDNA clone |
| KEYWORDS | | IMAGE:118649 5', mRNA sequence. |
| SOURCE | T92497.1 GI:724410 | |
| ORGANISM | human. | |
| | Homo sapiens | |
| | Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| REFERENCE | 1 (bases 1 to 337) | |
| AUTHORS | Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chitapelli, B., | |
| | Chisoso, S., Dietrich, N., Dubucque, T., Favello, A., Gish, W., Hawkins | |


```

yz3iel2.rl Soares_multiple_sclerosis_2NbHSP Homo sapiens cDNA
clone IMAGE:284686 5', mRNA sequence.
N73434
GI:1230719
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 161)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Ie,N., Wardis,E., Moore
,B., Morris,M., Parsons,J.J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1021 Std Error: 0.00
Seq primer: reverse RT
High quality sequence stop: 145.
FEATURES
source Location/Qualifiers
1..161
/organism="Homo sapiens"
/db_xref="GDB:3903214"
/db_xref="taxon:9606"
/clone="IMAGE:284686"
/clone_lib="Soares_multiple_sclerosis_2NbHSP"
/sex="male"
/tissue.type="multiple sclerosis lesions"
/dev_stage="Age 45"
/lab_host="DH10B (ampicillin resistant)"
/note=Vector: pT73D (Pharmacia) with a modified
polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TCCTACCAATCGAAGTCGGAGCCGCCGCATTTTTTTTTTTTTT 3'/],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Rento
Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 36 a 45 c 44 g 36 t
ORIGIN
Query Match 57.5%; Score 23; DB 10; Length 161;
Best Local Similarity 74.4%; Pred. No. 1.le+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 2 tcgagccaccataacctcaactctcaggattggg 40
||| ||||| || ||| ||| ||||| |||||
Db 91 TCACCACACACTTGACGCCACAACATCGCAGGGATTGGG 53
RESULT 10
BF090424
LOCUS BF090424 208 bp mRNA linear EST 19-OCT-2000
DEFINITION RC1-NT0033-110900-111-c09 NT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF090424
VERSION BF090424.1 GI:10896134

```

| | |
|------------|--|
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| REFERENCE | 1 (bases 1 to 208) |
| AUTHORS | Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare .M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. |
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) |
| MEDLINE | 20202663 |
| COMMENT | Contact: Simpson A J G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC1-NT0033-110 900-111-c09&t3=2000-09-11&t4=1) Seq primer: puc 18 forward High quality sequence stop: 208. Location/Qualifiers 1..208 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="NT0033" /dev_stage="Adult" /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from QRESPEC PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." |
| BASE COUNT | 53 a 62 c 51 g 42 t |
| ORIGIN | |
| | Query Match 57.5%; Score 23; DB 10; Length 208; |
| | Best Local Similarity 74.4%; Pred. No. 1.1e+02; |
| | Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0; |
| Qy- | 2 tcagccacccttaacctcaactcattcgaggatggg 40 |
| | |
| Db | 115 TCCAGCCACACTTGAGGCCACAACCTGCAGGGATTGTG 153 |
| RESULT 11 | |
| BE089733/c | |
| LOCUS | BE089733 223 bp mRNA linear EST 12-JUN-2000 |
| DEFINITION | QV0-BT0704-120500-224-c08 BT0704 Homo sapiens cdNA, mRNA sequence. |
| ACCESION | BE089733 |
| VERSION | BE089733.1 GI:8480168 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| REFERENCE | 1 (bases 1 to 223) |
| AUTHORS | Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare .M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. |

Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV0-BR0704-120)
500-224-c08t3-2000-05-12st4-1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 223.

FEATURES
source
1..223
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BR0704"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions." 54 t
BASE COUNT 47 a 52 c 70 g

Query Match 57.5%; Score 23; DB 9; Length 223;
Best Local Similarity 74.4%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccaccctaaacccccaataactcattccaggatggg 40
||| ||||| | | | | | | | | | | | | | |
Db 86 TCCACCCACACTTGAGGCCACAACTGCAGGGATTGGG 48

RESULT 12
BF340702/c

LOCUS
DEFINITION
BF340702 225 bp mRNA linear EST 22-NOV-2000
602035458f1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183263
BF340702
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 225)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAW499 row: d column: 16
High quality sequence stop: 225.

THIS PAGE BLANK (USPTO)


```

RESULT      3
US-09-115-934A-1/c
; Sequence 1, Application US/09115934A
; Patent No. 6130317
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Xu, Qunli
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115-934A

```

```
?
?   FILING DATE:
?
?   CLASSIFICATION: 435
?
?   PRIOR APPLICATION DATA:
?
?     APPLICATION NUMBER: US 08/818-514
?     FILING DATE: 14-MAR-1997
?     ATTORNEY/AGENT INFORMATION:
?       NAME: Campbell, Cathryn A.
?       REGISTRATION NUMBER: 31,815
?       REFERENCE/DOCKET NUMBER: P-LJ 3209
?     TELECOMMUNICATION INFORMATION:
?       TELEPHONE: (619) 535-9001
?       TELEFAX: (619) 535-8949
?     INFORMATION FOR SEQ ID NO: 2:
?       SEQUENCE CHARACTERISTICS:
?         LENGTH: 2634 base pairs
?         TYPE: nucleic acid
?         STRANDEDNESS: single
?         TOPOLOGY: linear
?   US-09-115-934A-2
```


Db 2326 AGCTGCCCTCACCTGGATCCACGAGGACAGG 2291

RESULT 10

US-09-221-235-10
; Sequence 10, Application US/09221235
; Patent No. 6043040

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,235
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-235-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctgagcaccaccataaccctcaataactccagggattggg 40

||||| ||||| ||||| ||||| ||||| |||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 11

US-09-221-928-10

; Sequence 10, Application US/09221928
; Patent No. 6121030

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,928
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-928-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctgagcaccaccataaccctcaataactccagggattggg 40

||||| ||||| ||||| ||||| ||||| |||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 12

US-09-221-527-10

; Sequence 10, Application US/09221527
; Patent No. 6146832

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,527
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-527-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctgagcaccaccataaccctcaataactccagggattggg 40

||||| ||||| ||||| ||||| ||||| |||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 13

US-09-221-236-10

; Sequence 10, Application US/09221236
; Patent No. 6146841

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,236
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-236-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctgagcaccaccataaccctcaataactccagggattggg 40

||||| ||||| ||||| ||||| ||||| |||||

Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 14

US-09-221-416-10

; Sequence 10, Application US/09221416
; Patent No. 6153417

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,416
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115

```
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-416-10

Query Match      48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ctcgagccaccataaccctcaatactaccagggtggg 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

RESULT 15
US-09-221-245-10
; Sequence 10, Application US/09221245
; Patent No. 6180358
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,245
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: US 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-245-10

Query Match      48.0%; Score 19.2; DB 4; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ctcgagccaccataaccctcaatactaccagggtggg 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994

Search completed: June 4, 2002, 17:47:09
Job time: 6582 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 15:57:52 ; Search time 521.5 Seconds
(without alignments)
131.690 Million cell updates/sec

Title: US-09-721-543A-12

Perfect score: 40

Sequence: 1 ctcgagccaccataaacctcaataactccaggtggg 40

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 40 | 100.0 | 40 | 22 | AA10599 |
| 2 | 23 | 57.5 | 2634 | 19 | AAV59067 |
| 3 | 23 | 57.5 | 2922 | 21 | AA18054 |
| 4 | 21.2 | 53.0 | 121 | 22 | ABA71669 |
| 5 | 21.2 | 53.0 | 121 | 22 | ABA37784 |
| 6 | 21.2 | 53.0 | 121 | 22 | AAK20013 |
| 7 | 21.2 | 53.0 | 121 | 22 | AAK46055 |
| 8 | 21.2 | 53.0 | 121 | 22 | AA151965 |
| 9 | 21.2 | 53.0 | 473 | 22 | ABA59145 |
| | | | | | DNA ligand #5 for |
| | | | | | Bax inhibitor BI-1 |
| | | | | | Lung cancer associ |
| | | | | | Human foetal liver |
| | | | | | Probe #16250 for g |
| | | | | | Human brain expres |
| | | | | | Human bone marrow |
| | | | | | Probe #20651 used |
| | | | | | Human foetal liver |

| | | | | | | |
|------|------|------|--------|----|----------|---------------------|
| c 10 | 21.2 | 53.0 | 473 | 22 | ABA27923 | Probe #6388 for ge |
| c 11 | 21.2 | 53.0 | 473 | 22 | AAK07330 | Human brain expres |
| c 12 | 21.2 | 53.0 | 473 | 22 | AAK33095 | Human bone marrow |
| c 13 | 21.2 | 53.0 | 473 | 22 | AAI38893 | Probe #7579 used t |
| c 14 | 20.8 | 52.0 | 837 | 20 | AAK37525 | Human secreted pro |
| c 15 | 20.8 | 52.0 | 3896 | 23 | ABL06092 | Drosophila melanog |
| c 16 | 20.6 | 51.5 | 366 | 22 | AAI87560 | Human polynucleoti |
| c 17 | 20.6 | 51.5 | 744 | 22 | AAI96641 | Human neuroblastom |
| c 18 | 20.6 | 51.5 | 1907 | 22 | AAI08058 | Human extracellular |
| c 19 | 20.4 | 51.0 | 5592 | 22 | AAF30935 | Spodoptera frugipe |
| c 20 | 20 | 50.0 | 897 | 23 | AAI91288 | DNA encoding novel |
| c 21 | 20 | 50.0 | 1378 | 23 | AAI81354 | DNA encoding novel |
| c 22 | 20 | 50.0 | 3364 | 19 | AAV07266 | Muscle creatine ki |
| c 23 | 20 | 50.0 | 16217 | 24 | ABL32625 | Human immune syste |
| c 24 | 19.8 | 49.5 | 2832 | 15 | AAQ62186 | Cyanobacteria repl |
| c 25 | 19.8 | 49.5 | 4809 | 15 | AAQ62185 | pA01 plasmid fragm |
| c 26 | 19.8 | 49.5 | 15196 | 22 | AAK73103 | Human immune/haema |
| c 27 | 19.8 | 49.5 | 15196 | 22 | AAK87548 | Human immune/haema |
| c 28 | 19.8 | 49.5 | 32249 | 22 | ABA20005 | Human nervous syst |
| c 29 | 19.8 | 49.5 | 32249 | 22 | ABA07406 | Human pancreatic c |
| c 30 | 19.8 | 49.5 | 32249 | 22 | AAK91137 | Human digestive sy |
| c 31 | 19.8 | 49.5 | 50442 | 22 | AAK73083 | Human immune/haema |
| c 32 | 19.8 | 49.5 | 50442 | 22 | AAK87551 | Human immune/haema |
| c 33 | 19.6 | 48.0 | 511 | 23 | AAI80311 | DNA encoding novel |
| c 34 | 19.6 | 49.0 | 791 | 22 | AAI91505 | Human polynucleoti |
| c 35 | 19.6 | 49.0 | 325791 | 22 | AAI43104 | Human Oestrogen re |
| c 36 | 19.4 | 48.5 | 282 | 21 | AAA43173 | Xenopus secreted e |
| c 37 | 19.4 | 48.5 | 476 | 22 | ABA57621 | Human foetal liver |
| c 38 | 19.4 | 48.5 | 476 | 22 | AAK05674 | Human brain expres |
| c 39 | 19.4 | 48.5 | 476 | 22 | AAK31288 | Human bone marrow |
| c 40 | 19.4 | 48.5 | 476 | 22 | AAI37184 | Probe #5870 used t |
| c 41 | 19.4 | 48.5 | 674 | 20 | AAK37371 | Human secreted pro |
| c 42 | 19.4 | 48.5 | 1554 | 20 | AAK16675 | Xenopus WA545 prot |
| c 43 | 19.4 | 48.5 | 1617 | 23 | AAI85886 | DNA encoding novel |
| c 44 | 19.4 | 48.5 | 12789 | 24 | AAI16876 | Human interleukin |
| c 45 | 19.4 | 48.5 | 349980 | 22 | AAH64956 | C glutamicum codin |

ALIGNMENTS

RESULT 1
AA10599
ID AA10599 standard; DNA; 40 BP.
XX
AC AA10599;
XX
DT 24-SEP-2001 (first entry)
XX
DE DNA ligand #5 for human cytomegalovirus.
XX
Ligand; human cytomegalovirus; HCMV infection; herpetic ulcer;
KW pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW herpesvirus infection; lesion; ds.
XX
OS Human cytomegalovirus.
OS Synthetic.
XX
PN WO200138341-A1.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32259.
XX
PR 24-NOV-1999; 99US-0167509.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Liu F, Wang J, Jiang H;
XX
DR WPI; 2001-367658/38.
XX
PT Polynucleotide ligands useful as anti-viral agents for the treatment of

PT herpesvirus infections e.g. primary or chronic infections with human
XX cytomegalovirus -

PS Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 10 A; 15 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcgagccaccataacccctcaatactccagggtggg 40
|||||
Db 1 ctcgagccaccataacccctcaatactccagggtggg 40

RESULT 2
AAV59067/c
ID AAV59067 standard; cDNA: 2634 BP.

XX AAV59067;

XX 02-FEB-1999 (first entry)

XX Bax inhibitor BI-1 cDNA.

XX Bax inhibitor; BI-1; human; apoptosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FT CDS 73..786
FT /*tag= a

XX WO9840397-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US05015.

XX 14-MAR-1997; 97US-0818514.

XX (BURN-) BURNHAM INST.

XX Reed JC, Xu Q;

XX WPI; 1998-531519/45.

XX P-PSDB; AAW73136.

XX Bax inhibitor proteins, BI-1 and BI-2 - useful e.g. to modulate
PT cellular apoptotic activity or identify agents altering BI-1 or BI-2
PT binding which can modulate apoptotic activity

XX Claim 2; Page 61-63; 80pp; English.

XX This cDNA clone codes for an inhibitor protein, termed BI-1 (see
CC AAW73136), of the pro-apoptotic protein Bax. Nucleic acids encoding
CC BI-1 and BI-2 (see AAV59068) were identified by suppression of
CC Bax-induced death of yeast cells transformed to express human Bax.
CC A human HepG2 cDNA library was used for library screening. The
CC invention provides vectors, optionally expression or viral vectors,
CC containing BI nucleic acids, and host cells containing these
CC vectors. The nucleic acids encoding BI-1/BI-2 can be used to

CC increase expression of these proteins in cells, or antisense
CC molecules prepared from them used to decrease expression. In
CC these ways, cellular apoptotic activity may be modulated (claimed).
CC The nucleic acids and complementary sequences are also useful as
CC probes to detect BI-encoding nucleic acid molecules in samples.

XX Sequence 2634 BP; 632 A; 626 C; 597 G; 779 T; 0 other;

Query Match 57.5%; Score 23; DB 19; Length 2634;
Best Local Similarity 74.4%; Pred. No. 4.1;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccataacccctcaatactccagggtggg 40
|||||
Db 2056 TCCACCCACTTGAAGCCACAAACTGCAGGGATTGG 2018

RESULT 3
AAFI8054/c
ID AAFI8054 standard; DNA; 2922 BP.

XX AAFI8054;

XX 14-MAR-2001 (first entry)

XX Lung cancer associated polynucleotide sequence SEQ ID 73.

XX Human: lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnary;
KW gastrointestinal; nephrotropic; antinefactive; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI; 2000-587514/55.

XX P-PSDB; AAB58178.

XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -

XX Claim 1; Page 530-551; 1425pp; English.

XX Polynucleotide sequences AAFI7982 - AAFI8424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnary; gastrointestinal
CC general; nephrotropic; antinefactive; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAFI8425 - AAFI8433 and

PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 25; SEQ ID No 20651; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 121 BP; 46 A; 20 C; 28 G; 27 T; 0 other;

Query Match 53.0%; Score 21.2; DB 22; Length 121;
Best Local Similarity 76.5%; Pred. No. 14;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7 ccacccataaccctcaataactccaggattggg 40
DB 38 CCTCTCTATAATCCTCATTACTCAAGGTAATGGG 5

RESULT 9
ABA59145/C
ID ABA59145 standard; DNA; 473 BP.
AC ABA59145;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #7450.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 7450; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7 ccacccataaccctcaataactccaggattggg 40
DB 347 CCTCTCTATAATCCTCATTACTCAAGGTAATGGG 314

RESULT 10
ABA27922/C
ID ABA27922 standard; DNA; 473 BP.
AC ABA27922;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #6388 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID NO 6388; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;


```
XX 09-AUG-2001.
PD
XX
XX PF 30-JAN-2001; 2001WO-US00663.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488897/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX PS Claim 25; SEQ ID No 7579; 654pp; English.
XX
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7 caaccataaccctcaataactccaggattggg 40
DB 347 CCTCTATATATCTTACTCAAGTAATGGG 314

RESULT 14
AAX37525/c
ID AAX37525 standard; cDNA; 837 BP.
XX
XX AC AAX37525;
XX
XX DT 06-JUL-1999 (first entry)
XX
XX DE Human secreted protein cDNA fragment containing gene 75.
XX
XX KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX immune system disease; autoimmune disease; hepatic disease; lymphoma;
XX renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
XX cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
XX pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO9918208-A1.
XX
XX PD 15-APR-1999.
XX
XX PF 01-OCT-1998; 98WO-US20775.
XX
XX PR 02-OCT-1997; 97US-0060884.
XX
XX PR 02-OCT-1997; 97US-0060833.
XX
XX PR 02-OCT-1997; 97US-0060836.
XX
XX PR 02-OCT-1997; 97US-0060837.

PR 02-OCT-1997; 97US-0060838.
PR 02-OCT-1997; 97US-0060839.
PR 02-OCT-1997; 97US-0060843.
PR 02-OCT-1997; 97US-0060862.
PR 02-OCT-1997; 97US-0060866.
PR 02-OCT-1997; 97US-0060874.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
XX DR WPI; 1999-264022/22.
XX
XX DR P-PSDB; AAY07926.
XX
XX PT New isolated human genes and the secreted polypeptides they encode
XX
XX PS Claim 1a; Page 264; 368pp; English.
XX
XX CC This invention describes novel isolated human genes and the secreted
XX proteins they encode. The products of the invention are useful for
XX preventing, treating or ameliorating medical conditions, e.g. by protein
XX or gene therapy. Also pathological conditions can be diagnosed by
XX determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 101 polynucleotides, based
XX on which tissues they are most highly expressed in, and include
XX developing products for the diagnosis or treatment of cancer, tumours,
XX neurodegenerative disorders, developmental abnormalities and fetal
XX deficiencies, blood disorders, leukemias, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,
XX transplant rejection, disorders involving osteoclasts such as
XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
XX infections and AIDS. The human secreted proteins of the invention are
XX represented in AAY07852-Y07993 and the encoding nucleic acids are
XX represented in AAX37451-X37552.
XX
XX SQ Sequence 837 BP; 171 A; 249 C; 187 G; 228 T; 2 other;

Query Match 52.0%; Score 20.8; DB 20; Length 837;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ctgcagccacccataaccctcaataactccaggattggg 40
DB 250 CTCAGGCAGCCCCAGAACACTCCAGACGCTGGAGATTGGG 211

RESULT 15
ABL06092
ID ABL06092 standard; cDNA; 3896 BP.
XX
XX AC ABL06092;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12758.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
```

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB61989.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12758; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3896 BP; 1147 A; 931 C; 883 G; 935 T; 0 other;

Query Match 52.0%; Score 20.8; DB 23; Length 3896;
Best Local Similarity 70.0%; Pred. NO. 38;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ctcgagccacccataaccctcaataactccagggttggg 40
||||| ||||| || || |||||
Db 3333 ctcgagccacccacacacaggcacacgctccacggttggg 3372

Search completed: June 4, 2002, 17:56:12
Job time: 7100 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 15:55:32 ; Search time 2137.99 Seconds
(without alignments)
391.518 Million cell updates/sec

Title: US-09-721-543A-12
Perfect score: 40
Sequence: 1 ctcagaccacccataaccctcaatactccaggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|----|----|-------------|
| ----- | | | | | | | |

| | | | | | | | |
|---|----|------|------|--------|----|-------------|--------------------|
| C | 1 | 23 | 57.5 | 1653 | 9 | AK026605 | AK026605 Homo sapi |
| C | 2 | 23 | 57.5 | 2600 | 9 | HSTEGT | X75861 H.sapiens T |
| C | 3 | 23 | 57.5 | 2609 | 9 | BC000916 | BC000916 Homo sapi |
| C | 4 | 23 | 57.5 | 2634 | 6 | AR058920 | AR058920 Sequence |
| C | 5 | 23 | 57.5 | 2634 | 6 | AR058921 | AR058921 Sequence |
| C | 6 | 23 | 57.5 | 2634 | 6 | AR112791 | AR112791 Sequence |
| C | 7 | 23 | 57.5 | 2634 | 6 | AR112792 | AR112792 Sequence |
| C | 8 | 23 | 57.5 | 197189 | 2 | AC019168 | AC019168 Homo sapi |
| C | 9 | 23 | 57.5 | 215867 | 2 | AC084037 | AC084037 Homo sapi |
| C | 10 | 23 | 57.5 | 225818 | 2 | AC090998 | AC090998 Homo sapi |
| C | 11 | 22.8 | 57.0 | 171185 | 2 | AP002424 | AP002424 Homo sapi |
| C | 12 | 22.8 | 57.0 | 177097 | 2 | AP001569 | AP001569 Homo sapi |
| C | 13 | 22.8 | 57.0 | 179726 | 9 | AC007052 | AC007052 Homo sapi |
| C | 14 | 22.8 | 57.0 | 182411 | 2 | AC090408 | AC090408 Homo sapi |
| C | 15 | 22.8 | 57.0 | 188357 | 9 | AL159996 | AL159996 Human DNA |
| C | 16 | 22.8 | 57.0 | 189329 | 2 | AC090230 | AC090230 Homo sapi |
| C | 17 | 22.8 | 57.0 | 196869 | 2 | AC087535 | AC087535 Homo sapi |
| C | 18 | 22.8 | 57.0 | 200774 | 2 | AP001592 | AP001592 Homo sapi |
| C | 19 | 22.4 | 56.0 | 137845 | 2 | AC010367 | AC010367 Homo sapi |
| C | 20 | 22.4 | 56.0 | 151514 | 9 | HS471M13 | Z97198 Human DNA s |
| C | 21 | 22 | 55.0 | 452 | 11 | G22123 | G22123 human STS W |
| C | 22 | 22 | 55.0 | 34887 | 3 | TBR012139 | AJ012199 Trypanoso |
| C | 23 | 22 | 55.0 | 76727 | 9 | HS821D11 | AL021453 Human DNA |
| C | 24 | 22 | 55.0 | 112022 | 9 | HS821D11 | AJ009611 Homo sapi |
| C | 25 | 22 | 55.0 | 177540 | 9 | AC006533 | AC006538 Homo sapi |
| C | 26 | 21.8 | 54.5 | 145607 | 2 | AC105341 | AC105341 Homo sapi |
| C | 27 | 21.8 | 54.5 | 152339 | 2 | AC083931 | AC083931 Homo sapi |
| C | 28 | 21.8 | 54.5 | 154165 | 2 | AC087620 | AC087620 Homo sapi |
| C | 29 | 21.8 | 54.5 | 154599 | 2 | AC016997 | AC016997 Homo sapi |
| C | 30 | 21.8 | 54.5 | 157021 | 2 | AC020991 | AC020991 Homo sapi |
| C | 31 | 21.6 | 54.0 | 36415 | 9 | AC093645 | AC093645 Homo sapi |
| C | 32 | 21.6 | 54.0 | 37470 | 9 | CNS00YVD | AL096810 Homo sapi |
| C | 33 | 21.6 | 54.0 | 104436 | 9 | HS611N7 | AL035663 Human DNA |
| C | 34 | 21.6 | 54.0 | 122952 | 2 | LMFLCHR12_4 | Continuation (5 of |
| C | 35 | 21.6 | 54.0 | 122952 | 2 | AL513483 | AL513483 Homo sapi |
| C | 36 | 21.6 | 54.0 | 173655 | 2 | OSJN00181 | AL662983 Oryza sat |
| C | 37 | 21.6 | 54.0 | 210442 | 2 | AL445199 | AL445199 Homo sapi |
| C | 38 | 21.6 | 54.0 | 226860 | 9 | AL359920 | AL359920 Human DNA |
| C | 39 | 21.4 | 53.5 | 54525 | 2 | AC103846 | AC103846 Homo sapi |
| C | 40 | 21.4 | 53.5 | 70993 | 2 | AC102199 | AC102199 Mus muscu |
| C | 41 | 21.4 | 53.5 | 161820 | 2 | AC095091 | AC095091 Rattus no |
| C | 42 | 21.4 | 53.5 | 217345 | 2 | AC104393 | AC104393 Homo sapi |
| C | 43 | 21.2 | 53.0 | 65824 | 2 | AC099899 | AC099899 Mus muscu |
| C | 44 | 21.2 | 53.0 | 70594 | 2 | AC015786 | AC015786 Homo sapi |
| C | 45 | 21.2 | 53.0 | 116252 | 9 | AL445228 | AL445228 Human DNA |

ALIGNMENTS

| | |
|------------|--|
| RESULT | 1 |
| LOCUS | AK026605 |
| DEFINITION | Homo sapiens CDNA: FLJ22952 fis, clone KAT09742. |
| ACCESSION | AK026605 |
| VERSION | AK026605.1 |
| KEYWORDS | oligo capping; fis (full insert sequence). |
| SOURCE | Homo sapiens signet-ring cell carcinoma cell_line:KATO III CDNA to mRNA, clone:lib:KAT clone:KAT09742. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | 1 (sites) Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. |
| TITLE | NEDO human cDNA sequencing project |
| JOURNAL | Unpublished (2000) |
| REFERENCE | 2 (bases 1 to 1653) Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio |
| JOURNAL | |

```
/translation="MNIFDRKINFDAALLKFHSITPSTQQHLKKVYASFALCMFVAAAG  
AYVHMVTHFIQAGLLSALGSLILMIWLMATPHSHETQKRLGLLAGFAFLTGVLGPGA  
LEFCIAVNPSILPTAFMGTMAMIFCTLSALYARRRSYFLGGILMSALSLLLLSSLG
```

LEFCIAVNPSILPTA

LEFCIAVNPSILPTA

Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa;
Mammalia; Eutheria; Primates;
Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 197189)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 197189)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7021812.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0161E16
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: plasmid; 8%
Chemistry: Dye-terminator ET; 92% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180505 bases at least Q40
Consensus quality: 186042 bases at least Q30
Consensus quality: 189445 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 194889; sum-of-contigs
Quality coverage: 3.69 in Q20 bases; agarose-fp
Quality coverage: 3.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1451: contig of 1451 bp in length
* 1452 1551: gap of unknown length
* 1552 2906: contig of 1355 bp in length
* 2907 3006: gap of unknown length
* 3007 4304: contig of 1298 bp in length
* 4305 4404: gap of unknown length
* 4405 5621: contig of 1217 bp in length
* 5622 5721: gap of unknown length
* 5722 8099: contig of 2378 bp in length
* 8100 8199: gap of unknown length
* 8200 10800: contig of 2601 bp in length
* 10801 10900: gap of unknown length
* 10901 14200: contig of 3300 bp in length
* 14201 14300: gap of unknown length
* 14301 17571: contig of 3271 bp in length
* 17572 17671: gap of unknown length
* 17672 21856: contig of 4185 bp in length
* 21857 21956: gap of unknown length
* 21957 25748: contig of 3792 bp in length
* 25749 25848: gap of unknown length
* 25849 28842: contig of 3994 bp in length
* 28843 29943: gap of unknown length
* 29944 35421: contig of 5479 bp in length
* 35422 35521: gap of unknown length
* 35522 40974: contig of 5453 bp in length
* 40975 41075: gap of unknown length
* 41076 48591: contig of 7517 bp in length
* 48592 48692: gap of unknown length
* 48693 58168: contig of 9477 bp in length
* 58169 58269: gap of unknown length
* 58270 68032: contig of 9764 bp in length
* 68033 77463: gap of unknown length
* 77464 93029: contig of 9331 bp in length

77464 77563: gap of unknown length
* 77564 93029: contig of 15466 bp in length
* 93030 93129: gap of unknown length
* 93130 107954: contig of 14825 bp in length
* 107955 108055: gap of unknown length
* 108056 121842: contig of 13788 bp in length
* 121843 121943: gap of unknown length
* 121944 137362: contig of 15420 bp in length
* 137363 137462: gap of unknown length
* 137463 155524: contig of 18062 bp in length
* 155525 155625: gap of unknown length
* 155626 173416: contig of 17792 bp in length
* 173417 173517: gap of unknown length
* 173518 197189: contig of 23673 bp in length.
FEATURES
Location/Qualifiers
1. 197189
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-161E16"
1. 1451
/note="assembly_name:Contig3"
1552. 2906
/note="assembly_name:Contig4"
3007. 4304
/note="assembly_name:Contig5"
4405. 5621
/note="assembly_name:Contig6"
5722. 8099
/note="assembly_name:Contig7"
8200. 10800
/note="assembly_name:Contig8"
10901. 14200
/note="assembly_name:Contig9"
14301. 17571
/note="assembly_name:Contig10"
17672. 21856
/note="assembly_name:Contig11"
21957. 25748
/note="assembly_name:Contig12"
25849. 29842
/note="assembly_name:Contig13"
29943. 35421
/note="assembly_name:Contig14"
35522. 40974
/note="assembly_name:Contig15"
41075. 48591
/note="assembly_name:Contig16"
48692. 58168
/note="assembly_name:Contig17"
58269. 68032
/note="assembly_name:Contig18"
68133. 77463
/note="assembly_name:Contig19"
77564. 93029
/note="assembly_name:Contig20"
93130. 107954
/note="assembly_name:Contig21"
clone_end:SP6
vector_side:right
108055. 121842
/note="assembly_name:Contig22"
121943. 137362
/note="assembly_name:Contig23"
137463. 155524
/note="assembly_name:Contig24"
155625. 173416
/note="assembly_name:Contig25"
173517. 197189
/note="assembly_name:Contig26"
BASE COUNT 51312 a 47093 c 47437 g 49028 t 2319 others
ORIGIN

```

Query Match      57.5%; Score 23; DB 2; Length 197189;
Best Local Similarity 74.4%; Pred. No. 27;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccaccataaccctcaactactcaggattggg 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 123952 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 123990

RESULT 9
AC084037 Homo sapiens chromosome 12q clone RP11-382G3, WORKING DRAFT
LOCUS DEFINITION
ACCESSION AC084037
VERSION 28 GI:18449688
KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 215867)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayelo,M., Banks,T.,
Barbata,J., Benton,J., Bimager,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Jacobson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalke,I., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 215867)
Worley,K.C.
Direct Submission
Submitted (11-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 31, 2002 this sequence version replaced gi:15809132.
----- Genome Center

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCHJ
Center clone name: RP11-382G3
----- Summary Statistics
Sequencing vector: Plasmid; M13; L08821
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.990329first call to
findPhrapList
Consensus quality: 223543 bases at least Q40
Consensus quality: 227846 bases at least Q30
Consensus quality: 229698 bases at least Q20
Estimated insert size: 217093; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 96329: contig of 96329 bp in length
* 96330 96429: gap of unknown length
* 96430 176860: contig of 80431 bp in length
* 176861 176960: gap of unknown length
* 176961 189494: contig of 12534 bp in length
* 189495 189595: gap of unknown length
* 189596 202514: contig of 12920 bp in length
* 202515 208888: contig of 6274 bp in length
* 208889 208988: gap of unknown length
* 208989 213216: contig of 4228 bp in length
* 213217 213616: gap of unknown length
* 213617 215867: contig of 2551 bp in length.
*
* FEATURES
* Location/Qualifiers
* 1..215867
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="12q"
* /clone="RP11-382G3"
*
* BASE COUNT 59187 a 51141 c 50650 g 54279 t 610 others
* ORIGIN
*
Query Match 57.5%; Score 23; DB 2; Length 215867;
Best Local Similarity 74.4%; Pred. No. 27;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccaccataaccctcaactactcaggattggg 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 142155 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 142193

RESULT 10
AC090998 Homo sapiens chromosome 12q clone RP11-161E16, WORKING DRAFT
LOCUS DEFINITION
ACCESSION AC090998
VERSION 4 GI:18449710
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens

```



```

misc_feature      /note="assembly_fragment"
167805..169750
/note="assembly_fragment"
169851..171185
/note="assembly_fragment"
171185..171185
BASE COUNT      52031 a 32086 c 33289 g 52078 t 1701 others
ORIGIN

Query Match      57.0% Score 22.8; DB 2; Length 171185;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctcgagccaccccaataaccctcaatactcagg 34
    ||| |||| ||||| ||||| ||||| |||||
Db 93406 cttctagcctcccatgcccctcattctccagtg 93439

RESULT 12
LOCUS      AP001569      177097 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-859C21 map 18q21, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION      AP001569
VERSION        AP001569.2 GI:8117403
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens DNA, clone:RP11-859C21.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 177097)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 177,097 genomic DNA of 18q21
PUBLISHED ONLY in DataBase (2000) In press
REFERENCE      2 (bases 1 to 177097)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
TITLE           Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@sc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT        On May 30, 2000 this sequence version replaced gi:7380904.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@sc.riken.go.jp
----- Project Information
Center project name: Humdraft18
Center clone name: RP11-859C21
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Ex-amerisham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 160670 bases at least Q40
Consensus quality: 168058 bases at least Q30
Consensus quality: 171631 bases at least Q20
Insert size: 173997; sum-of-contigs
Quality coverage: 4.63x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 23792 contig of 23792 bp in length
23893 45465 contig of 21573 bp in length

```

```

45566 58167 contig of 12602 bp in length
58268 71173 contig of 12906 bp in length
81253 91080 contig of 9980 bp in length
81354 91080 contig of 9727 bp in length
91181 97387 contig of 6207 bp in length
97488 104902 contig of 7415 bp in length
105003 111107 contig of 6105 bp in length
111208 116851 contig of 5644 bp in length
116952 123585 contig of 6634 bp in length
123686 128985 contig of 5300 bp in length
129086 134233 contig of 5148 bp in length
134334 138588 contig of 4255 bp in length
138689 142573 contig of 3885 bp in length
142674 146286 contig of 3613 bp in length
146387 150053 contig of 3667 bp in length
150154 154046 contig of 3893 bp in length
154147 156225 contig of 2079 bp in length
156326 158578 contig of 2253 bp in length
158679 161124 contig of 2446 bp in length
161225 163559 contig of 2335 bp in length
163660 165538 contig of 1879 bp in length
165639 166949 contig of 1311 bp in length
167050 168240 contig of 1191 bp in length
168341 169579 contig of 1239 bp in length
169680 170978 contig of 1299 bp in length
171079 172080 contig of 1002 bp in length
172181 173514 contig of 1334 bp in length
173615 174614 contig of 1000 bp in length
174715 175976 contig of 1262 bp in length
176077 177097 contig of 1021 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 23792: contig of 23792 bp in length
23893 23892: gap of 100 bp
23893 45465: contig of 21573 bp in length
45466 45565: gap of 100 bp
45566 58167: contig of 12602 bp in length
58168 58267: gap of 100 bp
58268 71173: contig of 12906 bp in length
71174 71273: gap of 100 bp
71274 81253: contig of 9980 bp in length
81254 81353: gap of 100 bp
81354 91080: contig of 9727 bp in length
91081 91180: gap of 100 bp
91181 97387: contig of 6207 bp in length
97388 97487: gap of 100 bp
97488 104902: contig of 7415 bp in length
104903 105002: gap of 100 bp
105003 111107: contig of 6105 bp in length
111108 111207: gap of 100 bp
111208 116851: contig of 5644 bp in length
116852 116951: gap of 100 bp
116952 123585: contig of 6634 bp in length
123586 123685: gap of 100 bp
123686 128985: contig of 5300 bp in length
128986 129085: gap of 100 bp
129086 134233: contig of 5148 bp in length
134234 134333: gap of 100 bp
134334 138588: contig of 4255 bp in length
138589 138688: gap of 100 bp
138689 142573: contig of 3885 bp in length
142574 142673: gap of 100 bp
142674 146286: contig of 3613 bp in length
146287 146386: gap of 100 bp
146387 150053: contig of 3667 bp in length
150054 150153: gap of 100 bp

```

* 150154 154046: contig of 3893 bp in length
* 154047 154146: gap of 100 bp
* 154147 156225: contig of 2079 bp in length
* 156226 156325: gap of 100 bp
* 156326 158578: contig of 2253 bp in length
* 158579 158678: gap of 100 bp
* 158679 161124: contig of 2446 bp in length
* 161125 161224: gap of 100 bp
* 161225 163559: contig of 2335 bp in length
* 163560 163659: gap of 100 bp
* 163660 165538: contig of 1879 bp in length
* 165539 165638: gap of 100 bp
* 165639 165649: contig of 1311 bp in length
* 165650 167049: gap of 100 bp
* 167050 168240: contig of 1191 bp in length
* 168241 168340: gap of 100 bp
* 168341 169579: contig of 1239 bp in length
* 169580 169679: gap of 100 bp
* 169680 170978: contig of 1299 bp in length
* 170979 171078: gap of 100 bp
* 171079 172080: contig of 1002 bp in length
* 172081 172180: gap of 100 bp
* 172181 173514: contig of 1334 bp in length
* 173515 173614: gap of 100 bp
* 173615 174614: contig of 1000 bp in length
* 174615 174714: gap of 100 bp
* 174715 175976: contig of 1262 bp in length
* 175977 176076: gap of 100 bp
* 176077 177097: contig of 1021 bp in length.

FEATURES

source

Location/Qualifiers
1. .177097
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="Rp11-859C21"
1. .23792
/note="assembly_fragment"
23893. .45465
/note="assembly_fragment"
45566. .58167
/note="assembly_fragment"
58268. .71173
/note="assembly_fragment"
71274. .81253
/note="assembly_fragment"
81354. .91080
/note="assembly_fragment"
91181. .97387
/note="assembly_fragment"
97488. .104902
/note="assembly_fragment clone_end:SP6 vector_side:left"
105003. .111107
/note="assembly_fragment"
111208. .116851
/note="assembly_fragment"
116952. .123585
/note="assembly_fragment"
123686. .128985
/note="assembly_fragment"
129086. .134233
/note="assembly_fragment"
134334. .138588
/note="assembly_fragment"
138689. .142573
/note="assembly_fragment"
142674. .146286
/note="assembly_fragment clone_end:T7 vector_side:left"
146387. .150053
/note="assembly_fragment"
150154. .154046
/note="assembly_fragment"
154147. .156225

/note="assembly_fragment"
Query Match 57.0%; Score 22.8; DB 2; Length 177097;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ctgcagccaccaccataaacctcctaatactcagg 34
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3709 CTGTAGCTCCCATGCCCTCATTTCTCCAGTG 3742

RESULT 13

AC007052/c

LOCUS Homo sapiens chromosome 18, clone HRPK.411_H_24, complete sequence.
AC007052
VERSION AC007052.4 GI:4510438
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179726)
AUTHORS Birten,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone HRPK.411_H_24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179726)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (09-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 179726)
REFERENCE 3 (bases 1 to 179726)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (23-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 24, 1999 this sequence version replaced gi:4432872.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html.
FEATURES
1. .179726
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"

```
repeat_region /map="18"
repeat_region /clone="hrpk.411.H_24"
repeat_region /clone_lib="RPCI-11 human BAC library"
repeat_region 898..1014
repeat_region /rpt_family="L1MD3"
repeat_region complement(1035..1333)
repeat_region /rpt_family="AluSx"
repeat_region 2987..3030
repeat_region /rpt_family="CT-rich"
repeat_region 3087..3291
repeat_region /rpt_family="L1MC4"
repeat_region 4298..4330
repeat_region /rpt_family="(CA)n"
repeat_region 5344..5365
repeat_region /rpt_family="AT-rich"
repeat_region complement(5552..5724)
repeat_region /rpt_family="AluJ"
repeat_region 5731..11847
repeat_region /rpt_family="L1PA2"
repeat_region 11857..11891
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(11893..12002)
repeat_region /rpt_family="AluJ"
repeat_region 11925..11948
repeat_region /note="Single-stranded coverage."
repeat_region 12026..12321
repeat_region /rpt_family="AluSg"
repeat_region complement(12605..12846)
repeat_region /rpt_family="MIR"
repeat_region 13352..13660
repeat_region /rpt_family="AluSc"
repeat_region complement(13741..14128)
repeat_region /rpt_family="MLT1B"
repeat_region 14425..14456
repeat_region /rpt_family="AT-rich"
repeat_region 14457..14978
repeat_region /rpt_family="MLT1P"
repeat_region 15496..15834
repeat_region /rpt_family="THE1B"
repeat_region 16325..16628
repeat_region /rpt_family="AluYb8"
repeat_region 16753..17086
repeat_region /rpt_family="MERT7A"
repeat_region 17543..17601
repeat_region /rpt_family="MIR"
repeat_region 18075..18442
repeat_region /rpt_family="MLT1B"
repeat_region complement(18735..18882)
repeat_region /rpt_family="MIR"
repeat_region complement(18948..19083)
repeat_region /rpt_family="L1MB8"
repeat_region complement(19190..19426)
repeat_region /rpt_family="AluJo"
repeat_region 19431..19750
repeat_region /rpt_family="L1ME3"
repeat_region 20434..20555
repeat_region /rpt_family="MLT1A1"
repeat_region 20556..20863
repeat_region /rpt_family="AluSx"
repeat_region 20864..21080
repeat_region /rpt_family="MLT1A1"
repeat_region 21528..21565
repeat_region /rpt_family="AT-rich"
repeat_region 21575..21668
repeat_region /rpt_family="L2"
repeat_region complement(21849..21991)
repeat_region /rpt_family="MER3B"
repeat_region 23147..23168
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(23169..23384)
repeat_region /rpt_family="AluJo"
repeat_region 23385..23419
repeat_region /rpt_family="AT-rich"
```

```
repeat_region complement(23799..23846)
repeat_region /rpt_family="L2"
repeat_region complement(23947..24060)
repeat_region /rpt_family="L2"
repeat_region 24066..24206
repeat_region /rpt_family="BC200"
repeat_region complement(24347..24514)
repeat_region /rpt_family="MER21B"
repeat_region 24782..25203
repeat_region /rpt_family="L1MB2"
repeat_region 25385..26005
repeat_region /rpt_family="L1MB2"
repeat_region 26035..26336
repeat_region /rpt_family="AluSx"
repeat_region 26432..26604
repeat_region /rpt_family="L1ME"
repeat_region 26610..26645
repeat_region /rpt_family="AT-rich"
repeat_region complement(27113..27422)
repeat_region /rpt_family="AluSp"
repeat_region 27865..28190
repeat_region /rpt_family="AluSg"
repeat_region 28250..28272
repeat_region /rpt_family="AT-rich"
repeat_region 28363..28496
repeat_region /rpt_family="HAL1"
repeat_region complement(28497..28796)
repeat_region /rpt_family="AluSg"
repeat_region 28797..28869
repeat_region /rpt_family="HAL1"
repeat_region 28993..29027
repeat_region /rpt_family="AT-rich"
repeat_region 29107..29134
repeat_region /rpt_family="(TATATG)n"
repeat_region 29151..29182
repeat_region /rpt_family="(TG)n"
repeat_region complement(29821..30164)
repeat_region /rpt_family="MER7A"
repeat_region complement(30236..30522)
repeat_region /rpt_family="AluJo"
repeat_region 30576..30723
repeat_region /rpt_family="MIR"
repeat_region 30927..31271
repeat_region /rpt_family="MLT1K"
repeat_region 32382..32437
repeat_region /rpt_family="(CA)n"
repeat_region 32480..32502
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(32503..32751)
repeat_region /rpt_family="AluSx"
repeat_region 34312..34476
repeat_region /rpt_family="MIR"
repeat_region complement(34648..34953)
repeat_region /rpt_family="AluSx"
repeat_region complement(35350..35947)
repeat_region /rpt_family="THE1B"
repeat_region 36048..36070
repeat_region /rpt_family="(TTTTA)n"
repeat_region complement(36072..36366)
repeat_region /rpt_family="AluJo"
repeat_region 36525..36587
repeat_region /rpt_family="AT-rich"
repeat_region 36792..36849
repeat_region /rpt_family="AT-rich"
repeat_region complement(36903..37092)
repeat_region /rpt_family="L1MA7"
repeat_region 37502..37812
```

Query Match

Best Local Similarity 57.0%; Score 22.8; DB 9; Length 179726;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctcgagccacccataacccctcaatctccagg 34

Qy 1 ctcgagccaccataaacctcaataactccagg 34
 ||| |||| ||||| ||||| ||||| ||||| |||||
 Db 122123 CTCAGCCTCCCATGCGCCCTCATTCCTCCAGTG 122090

RESULT 15

ALL159996 188357 bp DNA linear PRI 24-JUL-2000
 LOCUS Human DNA sequence from clone RP11-401F12 on chromosome 9. Contains
 DEFINITION STFS and GSSs, complete sequence.
 ACCESSION ALL159996
 VERSION ALL159996.7 GI:8648450
 KEYWORDS HYG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 188357)
 Laired.G.

REFERENCE Direct Submission
 AUTHORS Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
 JOURNAL CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

COMMENT On Jun 21, 2000 this sequence version replaced gi:8346239.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr9

RP11-401F12 is from the library RP11-11.2 constructed at the
 Roswell Park Cancer Institute by the group of Pieter de Jong. For
 further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-401F12 The true
 left end of clone RP11-208G24 is at 167549 in this sequence. The
 true right end of clone RP11-395D3 is at 5200 in this sequence.

FEATURES

source
 1..188357
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-401F12"
 /clone_lib="RP11-11.2"
 /clone 410..897
 misc_feature
 /note="match: GSS: Em:AQ0606829"
 1734..2348
 misc_feature
 /note="match: GSS: Em:AQ823967"
 1819..2460
 repeat_region
 /note="LIMC4 repeat: matches 6462..7077 of consensus"
 2551..2614
 repeat_region
 /note="2 copies 32 mer 96% conserved"
 2690..2698
 misc_feature
 /note="TN10 excised: This sequence represents the
 duplicated flanking sequence of the TN10."
 3529..3560
 repeat_region
 /note="8 copies 4 mer gtgt 96% conserved"
 3746..3959
 repeat_region
 /note="Alu repeat: matches 3..290 of consensus"
 3969..4006
 repeat_region
 /note="U2 repeat: matches 1..38 of consensus"
 4392..4430
 repeat_region
 /note="13 copies 3 mer aac 92% conserved"
 5217..5472
 misc_feature
 /note="match: GSS: Em:AQ302480"
 5260..6024
 misc_feature
 /note="match: GSS: Em:AQ901451"
 complement(5743..6091)
 misc_feature
 /note="match: GSS: Em:AQ132828"
 6101..6658
 misc_feature
 /note="match: GSS: Em:AQ830958"
 7246..8450
 repeat_region
 /note="Tigger3b repeat: matches 8..1231 of consensus"
 8626..8773
 repeat_region
 /note="L2 repeat: matches 2452..2611 of consensus"
 10391..10584
 repeat_region
 /note="MER5A repeat: matches 1..189 of consensus"
 11880..12174
 repeat_region
 /note="L2 repeat: matches 1986..2272 of consensus"
 complement(12504..12890)
 misc_feature
 /note="match: STS: Em:ISA311Y1"
 12609..12648
 repeat_region
 /note="10 copies 4 mer gtgt 90% conserved"
 12610..12647
 repeat_region
 /note="19 copies 2 mer tg 92% conserved"
 12615..12650
 repeat_region
 /note="6 copies 6 mer gtgtgt 91% conserved"
 13805..14099
 repeat_region
 /note="Alusg repeat: matches 2..295 of consensus"
 14101..14388
 repeat_region
 /note="Alusq repeat: matches 16..302 of consensus"
 15605..15839
 repeat_region
 /note="L2 repeat: matches 2357..2612 of consensus"
 15947..16013
 repeat_region
 /note="L2 repeat: matches 2609..2674 of consensus"
 16084..16364
 repeat_region
 /note="AluJo repeat: matches 1..288 of consensus"
 complement(16254..16783)
 misc_feature
 /note="match: GSS: Em:AQ764225"
 17436..17528
 repeat_region
 /note="MIR repeat: matches 35..123 of consensus"
 18177..18298
 repeat_region
 /note="FLAM-A repeat: matches 1..122 of consensus"
 19380..19675
 repeat_region
 /note="Alusx repeat: matches 1..296 of consensus"
 complement(19534..19956)
 misc_feature
 /note="match: GSS: Em:AQ042003"
 19736..19958
 repeat_region
 /note="L2 repeat: matches 2477..2745 of consensus"
 20126..20237
 repeat_region
 /note="L2 repeat: matches 2589..2705 of consensus"
 20737..20812
 repeat_region
 /note="MADE1 repeat: matches 5..80 of consensus"
 21025..21200
 repeat_region
 /note="L2 repeat: matches 2184..2355 of consensus"
 21443..21646
 repeat_region
 /note="MIR repeat: matches 40..246 of consensus"
 23474..23793
 repeat_region
 /note="MLTIF repeat: matches 197..541 of consensus"
 complement(25531..26230)
 misc_feature
 /note="match: GSS: Em:AQ787911"
 27777..28867
 repeat_region
 /note="LIMC3 repeat: matches 6667..7734 of consensus"
 28802..28880
 repeat_region
 /note="LIMC2 repeat: matches 6247..6326 of consensus"
 28924..29133
 repeat_region
 /note="MER20 repeat: matches 2..214 of consensus"
 29810..29849
 repeat_region
 /note="20 copies 2 mer ac 80% conserved"
 29868..30157
 repeat_region
 /note="Alu repeat: matches 3..290 of consensus"

Job time: 6609 sec

```
repeat_region 30822..31276
/note="MLT1H repeat: matches 70. .532 of consensus"
repeat_region 31665..32020
/note="THE1B repeat: matches 1. .364 of consensus"
misc_feature 32181..32674
/note="match: GSS: Em:AQ822292"
repeat_region 32279..32403
/note="MIR repeat: matches 57. .192 of consensus"
misc_feature complement(32857..33263)
/note="match: GSS: Em:AQ139531"
repeat_region 33279..33551
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 35349..35388
/note="10 copies 4 mer acac 92% conserved"
repeat_region 35353..35388
/note="3 copies 12 mer 94% conserved"
repeat_region 35759..36002
/note="MIR repeat: matches 2. .246 of consensus"
misc_feature 36104..36469
/note="match: GSS: Em:AQ031409"
repeat_region 37497..37545
/note="MIR repeat: matches 81. .127 of consensus"
repeat_region 37664..38248
/note="MLT2D repeat: matches 1. .553 of consensus"
repeat_region 38259..38373
/note="MIR repeat: matches 122. .249 of consensus"
repeat_region 38334..38392
/note="L2 repeat: matches 2692. .2750 of consensus"
misc_feature 38609..38867
/note="match: STS: Em:G13414"
repeat_region 38781..38991
/note="LTR33 repeat: matches 302. .518 of consensus"
repeat_region 39086..39212
/note="LTR33 repeat: matches 29. .165 of consensus"
repeat_region 39249..39483
/note="MER4D repeat: matches 1. .234 of consensus"
repeat_region 39483..40142
/note="MER4D repeat: matches 359. .1017 of consensus"
repeat_region 40325..40415
/note="MER94 repeat: matches 39. .134 of consensus"
repeat_region 41795..41902
/note="LTR33 repeat: matches 370. .472 of consensus"
repeat_region 41903..42192
/note="AluSg repeat: matches 1. .291 of consensus"
repeat_region 42193..42241
/note="LTR33 repeat: matches 472. .521 of consensus"
repeat_region 43494..43613
/note="L2 repeat: matches 2628. .2748 of consensus"
repeat_region 43695..43862
/note="FRAM repeat: matches 0. .164 of consensus"
repeat_region 44019..44082
/note="L2 repeat: matches 2356. .2428 of consensus"
repeat_region 44083..44384
/note="LIPM repeat: matches 5841. .6144 of consensus"
repeat_region 44385..44529
/note="L2 repeat: matches 2252. .2356 of consensus"
repeat_region 44541..44628
/note="MIR repeat: matches 128. .229 of consensus"
repeat_region 44631..44747
/note="MER33 repeat: matches 203. .324 of consensus"
repeat_region 44779..44985
```

```
Query Match 57.0%; Score 22.8; DB 9; Length 188357;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 ccaccccaataacctcaatactccagggtggg 40
   ||||| | | | | | | | | | | | | | |
Db 56526 CCACCCCATGCTCTCAAGACTCCAGGGAAGG 56559
```

Search completed: June 4, 2002, 17:45:41

THIS PAGE BLANK (USPTO)